PROTEOMICS OF *BACTEROIDES FRAGILIS* AND *ENTEROBACTER CANCEROGENUS*

A thesis submitted in partial fulfilment of the requirements of the University of Northumbria at Newcastle for the degree of Doctor of Philosophy

LAKSHMY MANICKAN

Research undertaken in the School of Applied Sciences

January 2010

Abstract

Bacteroides fragilis NCTC 9343 is a Gram-negative anaerobic bacterium with genomic DNA of 5205 Kb and a GC ratio of 43%. It is a commensal organism that can act as an opportunistic pathogen and is commonly present on the mucous membranes. It causes a variety of infections including intra abdominal infections, perirectal abscesses and decubitus ulcers. Enterotoxigenic forms are capable of causing diarrhoea in children and animals.

Enterobacter cancerogenus ATCC 35316 is also a Gram-negative facultatively anaerobic bacterium with genomic DNA of 4602 Kb and a GC ratio of 55%. It is a naturally occurring human gut symbiont known to exhibit resistance to antibiotics like aminopenicillins. It has also been reported in cases of severe osteomyelitis and infections of bones and joints.

This study aims to analyse the differential expression of proteins in the presence of mucin since it serves as the first site of adherence for the bacteria. The *E. cancerogenus* and *B. fragilis* proteins were extracted and separated by two dimensional electrophoresis from logarithmic phase cultures grown in semi-defined media enriched with or without porcine gastric mucin Types II and III. The gel images were analysed using Bio-Rad PDQuest, Ludesi Redfin and Nonlinear Dynamics SameSpots softwares. It was observed that the presence of mucin in the media affected the expression of a number of proteins in *E. cancerogenus* and *B. fragilis* cells. The protein spots of interest were excised, hydrolysed using trypsin and subjected to electrospray ionisation based LC-MS analysis in order to determine the identity of the digested proteins and obtain a better understanding of the interactions of *B. fragilis* and *E. cancerogenus* with mucin.

The outer membrane protein surface antigen X was found to be up-regulated in both mucin Type II and III enriched media in *E. cancerogenus*. Some of the other proteins that were differentially regulated in both *E. cancerogenus* and *B. fragilis* included the elongation factor Ts, malate dehydrogenase, triose phosphate isomerase and thiol peroxidase proteins indicating that these proteins may be associated with the ability of bacteria to grow in mucin and may be potential virulence factors.

Genes encoding the proteins CAH06598 and CAH09443 from the glycoside hydrolase families 95 and 97 in *B. fragilis* strain NCTC9343 were cloned, overexpressed and purified using nickel affinity and gel filtration chromatography. The enzymes were found to be active by performing fluorimetric assays using methyl-umbelliferyl sugar substrates. Diffracting crystals of CAH09443 were obtained from the PACT ANION screens containing polyethylene glycol and sodium malonate as a precipitant. Structure determination was achieved via molecular replacement using the glycoside hydrolase Family 97 α-galactosidase, *Bt*GH97b, from *Bacteroides thetaiotaomicron* as a starting model. The structure of CAH09443 was shown to be composed of a N-terminal β-super-sandwich domain and a canonical $(\beta/\alpha)_8$ barrel, similar to the two other glycoside hydrolase family 97 enzyme structures reported.

Acknowledgements

First of all, I would like to sincerely thank Prof. Gary Black for providing me with an opportunity to do this PhD by supporting and guiding me throughout the three years. The contributions and inputs provided by Dr. Ken Cook and Prof. Iain Sutcliffe are also very much appreciated. I express my gratitude to Dr. Edward Taylor at the University of York for all his help with the crystallisation work.

I would like to thank Dr. Meng Zhang and Dr. Anna Lindsay for helping me with learning the techniques used in the study. I would like to express my gratitude to all my lovely friends and fellow scientists in lab A307 with a special mention to Claire Jennings for putting up with my madness in the laboratory. The assistance provided by all the technicians in the School of Applied Sciences for performing experiments and using equipments is also appreciated. I would like to thank the IT Department for their support, especially Geoff Sparrow for downloading and installing all necessary software and the staff of Stephenson Building (Jeff Elliott and Dave Robson) for their indirect support and co-operation.

I would like to dedicate this thesis to my loving parents and sister, Lavanya without whose unconditional support, help and encouragement, this work would have been impossible. Finally, I thank Northumbria University for making this experience a special and memorable one.

Abbreviation list

APS	Ammonium persulphate
bp	Base pair
BSA	Bovine serum albumin
CBM	Carbohydrate binding module
CFE	Cell free extract
Da	Dalton
dNTP	deoxynucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
GH	Glycoside hydrolase
HEPES	N-[2-hydroxyethyl] piperazine-N'-[2-ethanosulphonic acid]
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilobase pair
LB	Luria Bertani
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
SDS	Sodium dodecyl sulphate
TAE	Tris-acetate-EDTA
TEMED	N,N,N',N'-tetramethylethylene diamine
TRIS	tris(hydroxymethyl)aminomethane
v/v	volume per volume

w/v	weight per volume
ACN	Acetonitrile
ATP	Adenosine tri phosphate
BHI	Brain Heart Infusion
BPB	Bromophenol blue
BPC	Base peak chromatogram
CHES	N-cyclohexyl-2-aminoethane sulphonic acid
DHAP	Dihydroxy acetone phosphate
ESI	Electrospray ionisation
EF	Elongation factor
IAA	Iodoacetamide
MES	2-(N-morpholino)-ethanesulphonic acid
MPD	2-Methyl-1,3 propanediol
MS	Mass spectrometry
FA	Formic acid
GDP	Guanosine di phosphate
G-3-P	Glyceraldehyde-3-phosphate
GTP	Guanosine tri phosphate
IEF	Isoelectric focussing
LC	Liquid chromatography
CAZY	Carbohydrate active enzymes
dCMP	deoxy cytidine mono phosphate
MPD	2-methyl-1,3 propanediol

NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NCBI	National Centre for Biotechnology Information
Omp	Outer membrane protein
PEG	Polyethylene glycol
PEG MME	Polyethylene glycol monomethyl ether
PEPCK	Phosphoenol pyruvate carboxykinase
PBS	Phosphate buffered saline
PFL	Pyruvate formate lyase
SOB	Super optimal broth
SOC	Super optimal culture
SSP	Standard spot numbers
TIC	Total ion current
TPI	Triose phosphate isomerase

Contents

Appendix

List of figures

List of tables

NB: * Tables 5 to 10 have been included in the Appendix section E.

DECLARATION

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work.

NAME: LAKSHMY MANICKAN

SIGNATURE:

DATE: 26th April 2010

4. General introduction

The study aims at understanding the interaction between gut commensal bacteria and mucin. The GI tract is colonised by a number of bacteria which are opportunistic pathogens that can cause infection when the host immune system is impaired. This work mainly focuses on the gut infection causing bacteria, *Enterobacter cancerogenus* strain ATCC 35316 and *Bacteroides fragilis* strain NCTC 9343.

4.1 Bacteroidaceae

Based on the comparative analysis of the 16S rRNA sequences, *Bacteroides* species have been classified under the family 'Bacteroidaceae' and phylum 'Bacteroidetes'. They play an important role in the maintenance of human and animal health by colonising the intestine and forming an indigenous flora of the colon. They have been known to colonise the distal regions of the stomach where the gastric secretions are diluted and the hindgut beginning from the distal regions of the small intestine with increasing numbers in the terminal illeum and colon. They help in the fermentation of carbohydrates in the colon where the fermentation activity is very high at the caecum. As the carbohydrate sources become limited, the activity is slowed down in the transverse colon and directed towards protein fermentation (Smith *et al*., 2006).

4.2 Introduction to *Bacteroides fragilis*

Bacteroides fragilis is classified as a Gram-negative anaerobe which is included in the family of Bacteroideceae. They commonly occur on mucous membranes and are responsible for endogenous infections. *Bacteroides* species cause more than 50% of anaerobic infections. They normally range from 0.5 - 0.6 µm in diameter to 1.5 - 4.5 µm in length and are found in nature as round ended bacilli (Fig. 1). They have a genome size of 5205 Kb with a guanine+cytosine (GC) ratio of 42% (Brook, 2006). They are opportunistic pathogens present in the gastrointestinal flora and cause a number of intra-abdominal infections.

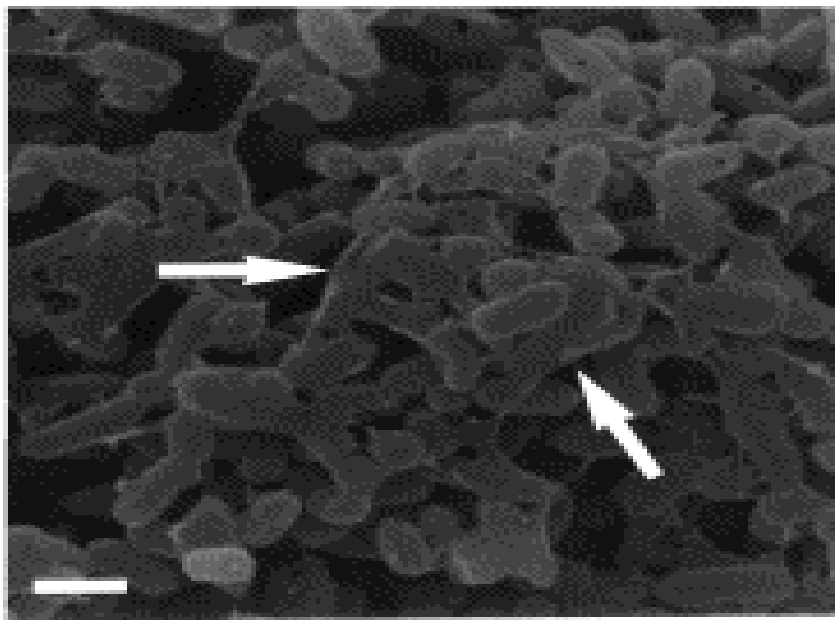


Figure 1: An electron micrograph showing aggregates of *B. fragilis.* The bar refers to a size of 1µm (Ferreiraa *et al*., 2002).

They have also been found to exhibit resistance to a variety of antibiotics including penicillin and the β-lactam antibiotics because they have the ability to produce the enzyme, β-lactamase which attacks the β-lactam ring within the antibiotic. They have also shown multidrug resistance to other antibiotics like clindamycin, tetracycline and metronidazole (Su & Honek, 2007).

4.2.1 Types of *Bacteroides fragilis* infections

B. fragilis is a Gram-negative bacterium that is capable of causing a variety of infections within the host. These include central nervous system infections, intra-abdominal and pleuropulmonary infections, skin and soft tissue infections, female genital tract infections, bacteraemia, osteomyelitis, abscesses and septic arthritis (Pumbwe *et al.*, 2006). *B. fragilis* is commonly present on the mucous membranes and causes a variety of infections including intra abdominal infections, perirectal abscesses and decubitus ulcers. The enterotoxigenic forms are capable of causing diarrhoea in children and animals (Nakano *et al.*, 2007).

4.2.2 Uses of *Bacteroides fragilis* to man

Bacteroides spp. are also beneficial to man in a number of ways. They are capable of producing certain vitamins like Vitamin K2 and menaquinone which are utilised by the human host.

Anaerobic growth of *Bacteroides* species results in the production of a number of end products such as butyrate, acetate and propionate and this contributes to nearly 70% of the energy source for enterocytes, a distinct type of epithelial cell that constitutes the innermost layer of the large and small intestine and helps to transport molecules into the tissues after breaking them down, present in the colon.

They play a very valuable role in the enterohepatic cycles which include the bile acid recirculation and the bile acid transformation by producing the enzyme, bile salt hydrolase (BSH) (Stellwag & Hylemon, 1976).

Bile acid recirculation involves the absorption of bile acids by the hepatocytes and this process is repeated several times within the intestine to make sure that the maximum amounts of nutrients are absorbed. The presence of anaerobic bacteria help to enhance this process (Hopley & Schalkwyk, 2006).

Bile acid transformation also plays an important role in the regulation of bile and cholesterol in the system. The terminal ileum and large bowel are largely characterised by the presence of those anaerobes which enhance the deconjugation process (Hopley & Schalkwyk, 2006). They compete with other pathogens in the colon thereby developing a competition at food supply and receptor sites by producing acidic end products like acetic acid and lactic acid that lowers the pH of the environment. This regulates the bacterial growth in the intestine.

4.2.3 General characteristics of *Bacteroides fragilis*

The selective medium which can be used to distinguish *B. fragilis* is Bile esculin agar which has the presence of oxgall, ferric ammonium sulphate (colour change indicator) and an antibiotic, gentamycin. Gentamycin and oxgall act as inhibitors of

other facultative anaerobes and Gram-negative anaerobes. Esculin is hydrolysed by *B. fragilis* to produce dextrose and esculetin and this reacts with the salt to produce black complexes which get accumulated around the colony (Livingston *et al*., 1978).

B. fragilis have the ability to exhibit haemolytic activity. Studies showed that the haemolytic activity diminished in the presence of oxygen or hydrogen peroxide and was controlled by the *hlyA* and *B* genes (Robertson *et al.*, 2006). This was confirmed when haemolytic activity was observed in blood agar plates containing *E. coli* transformed with the *hlyA* and *B* genes.

4.2.4 Cell wall structure of *Bacteroides fragilis*

B. fragilis has a distinct outer capsular layer made up of polysaccharides A and B (Fig. 2) (Pruzzo *et al*., 1989). The capsular layer confers resistance to phagocytosis, assists in adherence to the host and also evades complement mediated lysis (Pumbwe *et al*., 2006).

The second layer, termed as the lipopolysaccharide layer, projects out from the outer membrane of the lipid layer. This can be of varying thickness exhibiting a considerable level of toxicity and also functioning as an adhesive when contacted by host cells.

They have the presence of fibrils which act as short fine appendages whose role in virulence is not yet determined even though it is suspected to be related to adhesion and biofilm formation. The next important component of the cell wall includes the outer membrane vesicles. They are known to produce endotoxins like neuraminidase, sialidase and fibrinogenolytic enzymes (Pumbwe *et al*., 2006). The fibrinogenolytic enzymes, chondroitin sulphatases and hyaluronidases are histolytic in nature and *B. fragilis* is also capable of producing antimicrobial chemicals, namely bacteriocins. Neuraminidase exhibits a hydrolytic activity on the mucin polysaccharide. One of the previous studies showed that neuraminidase activity plays a very important role in the successful survival of *Bacteroides* in mammalian

tissues thereby acting as one of the major virulence conferring antigens (Godoy *et al.*, 1993).

Studies suggest that *B. fragilis* produces outer membrane proteins which play an important role in maintaining the structure of the cell and the production of these proteins are controlled by a set of 4 *omp* genes (Wexler, 2002).

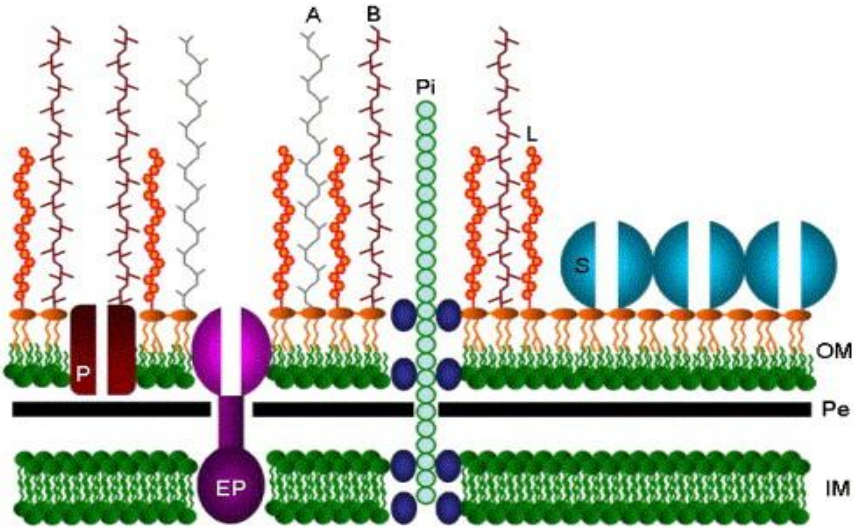


Figure 2:Cell wall structure of *B. fragilis* (Pumbwe *et al.,* 2006)

Pe- periplasmic space, OM- outer membrane, L- side chains, A and

B- Capsular polysaccharides, Pi- Pilus, P- outer membrane porin,

EP- efflux pump, IM- inner membrane.

4.2.5 Types of *B. fragilis* toxins

B. fragilis produces two main types of toxins,

- The endotoxin from the outer membrane vesicles and
- The enterotoxin whose production is controlled by the regulation of three *bft* genes, *I*, *II* and *III*. They can cause diarrhoea within the host by damaging the tight junctions of the intestine. Actin cytoskeleton rearrangement of the

epithelial cells causes a leakage in the internal contents of the intestine (Nakano *et al.*, 2007).

- The *bft* genes also induce a fluid secretion in the intestinal epithelia (Pumbwe *et al*., 2006).
- *B. fragilis* also produces proteins termed as bacteriocins which inhibit the growth of other bacteria by targeting their RNA polymerase thereby providing the bacteria a better chance of survival during competition from other pathogens (Stein, 1998).

4.2.6 *Bacteroides fragilis* virulence factors

Some of the virulence factors produced by *B. fragilis* include capsules where capsulated strains have been shown to be more virulent in nature (Pruzzo *et al*., 1989). They produce several extracellular or membrane bound enzymes like collagenases, fibrinolysin, proteases, lipases, ribonucleases and deoxyribonucleases involved in tissue degradation and destruction. The other virulence factors are the oxygen scavenging enzymes namely superoxide dismutases, catalases and peroxidases which enable the bacteria to survive in the presence of oxygen or in aerobic environments (Namavar *et al*., 1991). They produce endotoxins that are transported to the target host sites and induce complement activation. The capsular polysaccharides inhibit phagocytosis and induce abscess formation, the charged surface ligands help in adherence and bacterial interaggregation sometimes promoting the formation of biofilms of the same or different bacterial communities. The IgA protease causes impairment of secretory and mucosal immunity and heparinase promotes coagulation with tissue ischemia. Some strains of bacteria are capable of producing β-lactamases which confer resistance to β lactam antibiotics. Bacteriocins and other secondary metabolites including fatty acids, hydrogen sulphide and ammonia competitively inhibit the growth of normal flora (Stein, 1998).

Studies in *Bacteroides forsythus* have revealed the presence of a S-layer which is a unique surface structure that has been known to be involved in virulence. They help in the adherence of the bacterial cells to the host, surface recognition and resistance to phagocytosis. The S-layer ranges in size from 40-200 kDa in size and is composed of regularly aligned glycoprotein subunits. These proteins could contribute

to the ability of the bacteria to adhere and form biofilms on the surface of teeth resulting in dental plaque (Sabet *et al.*, 2003).

B. fragilis can exist in both capsulated and non-capsulated forms but the capsulated forms tend to have a higher degree of virulence conferring capacity. Their capsules contain two different types of polysaccharides called A and B which are linked together by oppositely charged groups and they confer resistance to the organism against the host defences (Pantosti *et al.,* 1993).

They also produce a number of virulence conferring enzymes which include neuraminidase, hyaluronate lyase, DNase, phosphatase and rare fibrinolysins. They produce the enzyme heparinase which may result in vascular thrombosis (Hopley & Schalkwyk, 2006). They also produce a toxin called fragilysin that is a zinc metalloprotease, 20 kDa in molecular weight. This damages the tight junctions of the intestine resulting in the destruction of enterocytes due to its cytotoxic effects and fluid secretion (Fasano & Nataro, 2004).

4.2.7 Treatment of *Bacteroides fragilis* infections

A number of normal antibiotics can be used for the treatment of *B. fragilis* infections like metronidazole, chloramphenicol, moxalactam, ceftriaxone, clarithromycin, cefoperazone, cefotaxine, ceftazidime, sparfloxacin, carbapenems and clindamycin (Katsandri *et al.*, 2006). However, recent studies showed that certain strains of *B. fragilis* exhibit resistance to carbapenem by producing a β-lactamase enzyme, carbapenemase. This was found to be induced by the activation of the *cfiA* gene with or without the insertion sequence elements indicating that this gene may be involved in the development of antibiotic resistance in *Bacteroides* (Edwards & Read, 2000). They also exhibited a resistance percentage of 0.8 and 1.3 for antibiotics imipenem and meropenem, respectively (Edwards & Read, 2000).

4.3 Enterobacteriaceae

Enterobacter species are capable of causing a number of infections ranging from bacteraemia, lower respiratory tract infections, skin and soft tissue infections to endocarditis, urinary tract infections, intra abdominal infections and post surgical wound infections. One of the major characteristic of Gram negative bacteria is their ability to produce lipopolysaccharides. Enterobacteriaceae produce lipopolysaccharides which are composed of three parts namely the oligomeric repeating O polysaccharide units, a core and lipid A. The core links the O polysaccharide units to the lipid A and this causes the lipopolysaccharide to remain anchored to the cell envelope. O polysaccharides exhibit structural variation which induces inflammatory responses in the host. The toxin is also capable of causing sepsis within the host which later leads to the development of *Enterobacter* bacteraemia characterised by cyanosis (Janda & Abbott, 2006). Biochemical characteristics reveal that these bacteria are straight rods, Gram negative, facultatively anaerobic and exhibit both respiratory and fermentative type of metabolism. They are citrate positive, indole negative and are chemoorganotrophic in nature (Bergey & Holt, 1994).

4.3.1 General virulence factors of *Enterobacter* species

Some of the major virulence factors associated with *Enterobacter* species are the mannose sensitive haemagglutinins (MSHA) type 1 or 3, siderophores, toxins and outer membrane proteins. The MSHA are putative fimbriae of 35 kDa size and have a receptor site that is recognised as a high mannose oligosaccharide. Siderophores have been known to be associated with invasive systemic infections. *E. cloacae* produces hydroxamate siderophore aerobactins that may be associated with invasive diseases whereas *E. coli* has been known to produce a 72 kDa aerobactin associated protein. Certain non-aerobactin hydroxamate compounds have also been found to be infrequently expressed in *Enterobacter* species. Details about their role in pathogenicity remain unknown (Keller *et al.*, 1998). *Enterobacter* species also produce toxins like α- haemolysin which have a glycine rich motif and exhibit cytotoxic activity against human erythrocytes and leukocytes. Molecular characterisation of invasive infections caused by *E. cloacae* showed an up-regulation

of the outer membrane protein OmpX suggesting that this could be a potential virulence factor, 17 kDa in size. It was observed that an up-regulation of *ompX* was accompanied by a down-regulation of other outer membrane porins OmpF and OmpC and may be associated with the ability of the organism to exhibit resistance to β-lactam compounds. Bacterial strains lacking the omp X gene were constructed to study their role in pathogenicity and it was observed that these proteins increased the infection rate by 10 fold in non mutant strains (Janda & Abbott, 2006).

4.3.2 *Enterobacter cancerogenus*

Enterobacter cancerogenus is also a Gram-negative facultatively anaerobic bacterium that is a naturally occurring human gut symbiont. They have a GC ratio of 55% with a size of 4602 Kb and are known to exhibit resistance to antibiotics like aminopenicillins. They have also been reported in cases of severe osteomyelitis and infection of bones and joints and belong to the CDC Enteric group 19.

This bacterium was formerly known as *Erwinia cancerogena* and several morphological, physiological and biochemical studies revealed that *E. cancerogena* was not a member of the genus *Erwinia* but was included in the genus *Enterobacter*. *E. cancerogenus* are capsulated, motile straight rod shaped bacteria with peritrichous flagella. *E. cancerogena* was found to exhibit 75% phenotypic similarity to *Erwinia carotovora* and 92% phenotypic similarity to *Enterobacter nimipressuralis*. Hence it was proposed that *E. cancerogena* could be moved to the genus *Enterobacter* based on phenotypic characterisation.(Dickey & Zumoff, 1988).

They are commonly found in nature and occur as commensals but when the host immune system is impaired they are capable of causing infections which make them opportunistic pathogens. Five cases of infection were reported in 1997. *E. cancerogenus* was identified from blood culture tests where wound infections and septicaemia occurred following environmental exposure of wounds from traumatic events or injuries in adults (Abbott & Janda, 1997).

Studies revealed that *E. cancerogenus* was formerly known as *Enterobacter taylorae* after being moved from the genus *Erwinia*. Based on the evidence obtained from the study of the biochemical and physiological properties of *E. taylorae*, it was proposed

that the organism needed to be included in the *Enterobacteriaceae* family. They were able to ferment glucose to produce gas and showed 84- 95% similarity to ATCC 35317 strain (Farmer *et al.*, 1985).There was a very strong relation between these two bacteria where they had a high degree of DNA relatedness. Since no difference could be ascertained between these two species, it was accepted that *E. cancerogenus* was a senior synonym of *E. taylorae* (Grimont & Ageron, 1989).

Even though it is known that these bacteria are capable of causing infections and occur predominantly as gut microbiota, details regarding their virulence factors or mode of action are not available. Several cases of infection have been reported since 1987 and the bacteria have been included in the Proteobacteria division and isolated mainly from human faeces. Infections occurred following a traumatic injury and the bacterium was found to exhibit resistance to aminopenicillins (Garazzino *et al.*, 2005). Four cases of nosocomial infections were reported in 1989 where all the cases described localised infection following open fractures, wounds or abrasions. The cases reported were bacteraemia, pneumonia and urinary tract infections. The bacteria were also found to produce β-lactamases showing resistance to antibiotics like penicillin and cephalosporins. Administration of cephalosporin prophylactically during cardiac surgery induced a drastic increase in the number of *Enterobacter* species in the intestine of the patient (Rubinstien *et al.*, 1993). Another reported case of *E. taylorae* infection was osteomyelitis in a patient following an open fracture wound being infected (Westblom & Coggins, 1987). The most recent occurrence of *E. cancerogenus* infection was reported in 2005 where the infection occurred following a multiple bone fracture with abundant environmental exposure (Garazzino *et al.*, 2005).

4.3.3 *Enterobacter* species attachment to intestinal cells

The outer membrane protein A has been known to play an important role in the attachment of *E. sakazakii* to the epithelial cells of the intestine of human hosts according to previous studies (Nair & Venkitanarayanan, 2007). Even though no reports are available regarding the attachment of *E. cancerogenus* to intestinal epithelial cells, *E. cloacae* which belongs to the same genus has been known to be associated with host intestinal tissues and mucin (Schierack *et al.*, 2007).

4.3.4 Antibiotic treatment

The commonly used antibiotic treatment for *Enterobacter* infections includes carbapenems and cefepime. Cross transmission contributes to some of the major factors that result in the spread of the bacteria. They have been predominantly found in intensive care units especially if infection control is poor.

4.4 Epithelial cells of the intestine and their role in protection against antigens

There are three types of cells that are produced to confer resistance to bacteria. This includes the goblet cells that produce proteoglycans and glycoproteins in the form of viscous mucin that covers the entire surface of the wet intestinal epithelial layer; the M cells that help to transport antigens from the exterior into the lymphoid tissue. This assists the host in identifying bacterial antigens and preparing the immune system. The third type of cells includes the Paneth cells that produce intra epithelial lymphocytes and antibacterial proteins that act against exogenous bacteria (Wilson, 2002). Intestinal cells also protect themselves against bacteria by shedding off the surface epithelia that fill up with keratin. This sloughing off of the outer layers results in the removal of adhered bacteria which get flushed away. Mucosal epithelia also have the ability to produce proteins that recognise lipopolysaccharides present exogenously. This may be produced due to continual exposure to certain specific surface components of bacteria or for identification of potential antigens (Laux *et al*., 2005).

The entire length of the gastrointestinal tract consists of the intestinal epithelial cells with the outer protective mucous membrane. The upper part of the tract has been known to be colonised by aerobic bacterial flora since the oxygen concentration of the intestine decreases from the duodenum to the colon. The lower part of the intestinal tract is colonised by the anaerobic flora. Variations in pH and nutrition can affect the microflora balance and also determines the general metabolism of the body. Changes in microflora of the gastrointestinal tract have been known to result in metabolic disorders (Serino *et al.*, 2009).

4.5 Mucin

Mucin is a heavily glycosylated protein with a high molecular weight. Mucin forms a major component of the body secretions which include the mucous and secretions of the mucous gland, the saliva and secretions of the salivary gland, fibres of the tendons and the connective tissue. It is albuminoid in nature with a ropy appearance and possesses heavy glycosylation (Dekker *et al*., 2002). Mucins contain hydrophobic membrane spanning domains that enable them to be retained within the plasma membrane making them membrane bound.

They generally form aggregates of proteins which are about 1-10 million Daltons and are known to contain neuraminic acids in their side chains (Dekker *et al*., 2002). They have the ability to resist proteolysis and act as the first line of defence against infection.

4.5.1 Structure of mucin

Mucin consists of a number of monomeric units that are linked by O and N linked oligosaccharides. Fig. 3 represents the mucin monomers linked together in an oligomeric gel and Figs. 4 and 5 represent the N and C termini of individual mucin monomers with N and O linked oligosaccharide units, cysteine knot and disulphide rich domains containing disulphide bonds. The disulphide rich domain is also termed as the D domain (Refer to Fig. 5) (Wilson, 2002). Porcine gastric mucin is predominantly made up of O-linked oligosaccharides and mucin side chains contain a number of groups attached to them including sulphates, neuraminic acids, sialic acids and so on.

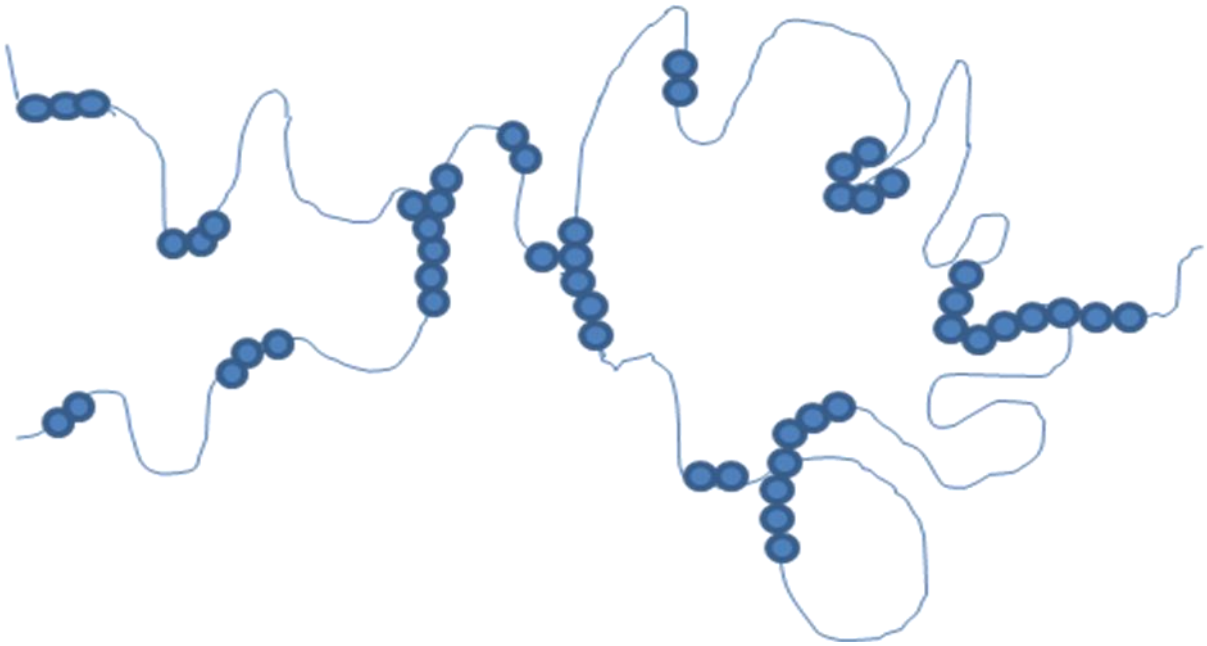


Figure 3: A number of mucin monomers linked together in an oligomeric gel

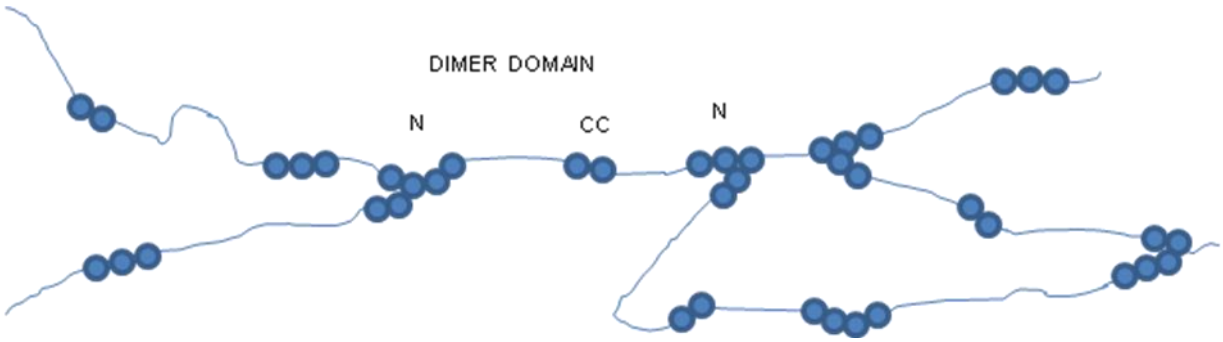


Figure 4: Structure of mucin showing the dimer domain

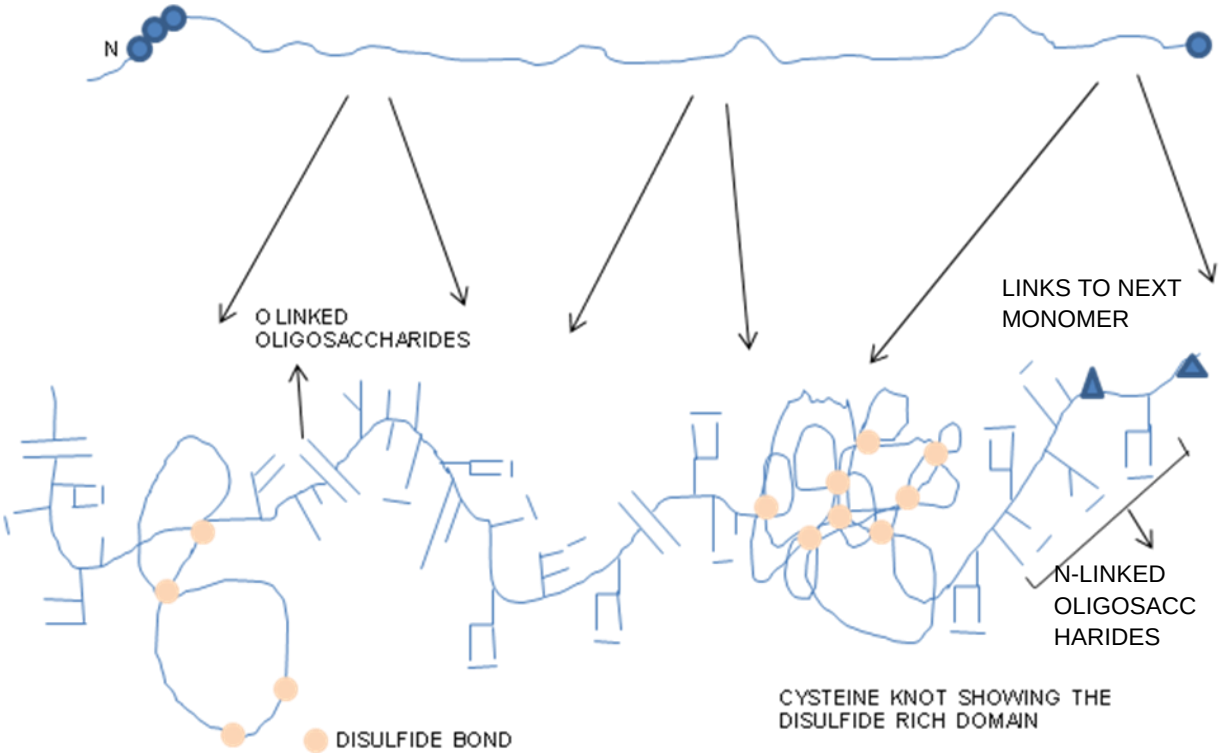


Figure 5: Oligomeric structure of mucin focussing on the monomer linkages and bonds (Wilson, 2002).

Mucins are of two main types namely soluble mucins and membrane mucins. Soluble mucin secretion associated with the respiratory tract is known to be encoded by a cluster of 18 genes that results in the formation of heavily glycosylated disulphide bond rich glycoproteins in the mucous in man (Rose & Voynow, 2006). Membrane mucins are highly diverse structurally and functionally and contain a transmembrane domain that enables them to associate with the cell surface thereby acting as the first line of defense and can mediate transport of solutes. Examples of membrane mucin include the mucin encoded by the muc1 and muc4 genes. They function as anti-adhesive agents that prevent cell-cell, cell-molecular interactions. They also act as modulators of cell signalling pathways where muc1 encodes for mucin that act as docking protein and muc4 encodes mucin that acts as a modulator of receptor tyrosine kinase ErbB2 (Zhang *et al.*, 2006). Sialomucin complexes have the ability to act as anti-adhesive agents by disrupting cell-matrix interactions and have been overexpressed in cancer cells (Komatsu *et al.*, 1997). This demonstrates the role of mucin and its protective functions within the host.

A study was conducted to study the variation of mucin distribution in the gut. The four main parts that were studied were the small intestine, caecum, colon and luminal contents and three types of mucin were found to predominantly occur in them. These were classified as neutral mucins, sialomucins and sulphomucins of which sialo and sulphomucins were found to be acidic in nature and sulphomucins were sulphated. It was found that conventional rats produced higher amounts to mucin with higher villi and deeper crypts in the small intestine when compared to germ free rats indicating that exposure to bacteria may be one of the main reasons for the production of mucin. It was also observed that conventional rats had decreased amounts of neutral and sulpho mucins in the intestine indicating that bacteria could have a higher hydrolytic activity on them when compared to sialomucins that exhibit higher resistance (Meslin *et al*., 1999). This could explain the presence of bound sialic acids in the partially purified and crude extracts of mucin from the porcine intestine.

The synthesis of mucin is controlled by a set of mucin genes termed *muc*. Mucin monomers are linked to form aggregates. They have specific non-covalent and intermolecular disulfide bonds to link them. The central region of the gene encodes the synthesis of hundreds of *O*-linked oligosaccharides of varying lengths whereas the N and C terminals have a much lesser level of glycosylation. The central region also possesses 10-80 repeating units of serine or threonine amino acid residues as shown in the figure below (Salyers *et al.*, 1977).

Figure 6: Structure of the mucin gene '*muc*'

4.5.2 Sugar composition of mucin

Some of the oligosaccharides which have been released upon breakdown of mucin include 2-acetamido-2-deoxy-α-D-galactopyranosyl (1→3)-[α-L-fucopyranosyl-(1→2)]-deoxy-β-D-galactopyranosyl-(1→3)-[N-glycoylneuraminyl-(2→6)]-2-acetamido-2-deoxy-D-galacitol, N-glycolylneuraminyl→N-acetylgalactosaminitol, N-acetylgalactosaminitol. Other oligosaccharides include GlcNAcβ(1-3)GalNAc-ol, Galβ (1-4)GlcNAcβ (1-3)GalNAc-ol, Galβ (1-4)GlcNAcβ (l-3)4Galβ (l-4)GlcNAcβ (1-3)GalNAc-ol, Galβ (1-4)GlcNAcβ (1-3)GalNAc-ol, GalNAcα(l-3)Galβ (1-4)GlcNAcβ (l-6), and GalNAcα(l-3)Galβ (l-4)GlcNAcβ (1-3)Galβ (l-4)GlcNAcβ (l-3)Galα(l-4)GlcNAcβ (l-3)GalNAc-ol irrespective of the source of mucin (Podolsky, 1985).

The saccharide composition of mucin depends upon the part of the body where it might be produced, nutrition and several other factors including the environmental conditions. Studies were performed in porcine submaxillary mucins based on their ability to inhibit human A- anti A haemagglutination. The breakdown of the complex polysaccharide into oligosaccharides was studied and the resulting oligosaccharides were grouped from I to V. The study aimed at determining the type of mucin oligosaccharides that was found to be active in inhibiting haemagglutination (Carlson, 1968). Breakdown of mucin resulted in the release of the monosaccharide, 2-acetamido-2-deoxy-D-galacitol (N-acetylgalactosaminitol), a pentasaccharide, 2-acetamido-2-deoxy-α-D-galactopyranosyl (1→3)-[α-L-fucopyranosyl-(1→2)]-deoxy-β-D-galactopyranosyl-(1→3)-[N-glycoylneuraminyl-(2→6)]-2-acetamido-2-deoxy-D-galacitol. The other reduced oligosaccharides released include N-acetylglucosamine, N-glycolylneuraminyl→N-acetylgalactosaminitol and N-acetylgalactosaminitol (Carlson, 1968).

Similar studies were performed in cervical mucin and it was found to contain neutral, sialylated and sulphated oligosaccharides including fucose, galactose, N-acetylgalactosamine, N-acetylglucosamine, N-acetyl-galactosamine and so on. A neutral tetrasaccharide and an acidic trisaccharide were also isolated. It has been observed that there is a high degree of heterogeneity in mucins derived from various parts of the body like bronchial mucin, colonic mucin, cervical mucin, gastric mucin and so on (Yurewicz & Moghissi, 1981).

Previous studies performed in colonic mucin showed the release of sugars like galactosamine, galactose, glucosamine, sialic acids, fucose, mannose and glucose (Podolsky & Isselbacher, 1983).

4.5.3 Functions of mucin

Mucins are present in various parts of the body and play a very important role by acting as protective barriers of the host against pathogenic bacteria. Intestinal mucin helps the system internally by protecting the walls against the acvtivity of strong digestive enzymes. They lubricate the walls of the intestine, exhibit surface hydrophobicity and can reduce the absorption of toxins (Lugea *et al.*, 2000). Innate mucosal defense involves two main components; mechanisms that are involved in preventing the invasion of bacteria and their toxins and the second mechanism involved the rapid regeneration of the mucosal layer by repairing damages and defects. They work in conjunction with the tight junctions of the intestine controlling the passive diffusion of solutes beneficial to the host while regulating the entry of bacterial toxins and antigens (Blikslager *et al.*, 2007). Even though it is known that the epithelial cell layer of the intestine secretes the mucin glycoproteins that form a viscoelastic gel, functional evidence of their protective functions are limited and not available in detail. Recent studies have revealed the presence of a distinctive and highly conserved structural motif containing 3 small proteins called the trefoil peptides. The goblet cells secreted these proteins and they were found to span the entire length of the gastrointestinal tract. The three proteins were designated as PS2, SP and the Intestinal trefoil factor (ITF) of which the first two were found to occur predominantly in the mucosa of the stomach, mouth, hepatobiliary duct and the pancreatic duct whereas the intestinal trefoil factor was found to be predominant in the small and large intestinal mucosa. These proteins contain a highly conserved 6 cysteine residue motif that form three intrachain loops with disulphide bond interactions. It has also been suggested that the occurrence of secondary structures of these proteins on the mucosal layer could confer resistance to the digestive activity of a variety of protease enzymes. This helps to maintain the intestinal walls structurally and functionally intact from the activity of proteases (Podolsky, 1999).

4.5.4 Mucin degradation by bacteria

Bacteria have the ability to produce a variety of enzymes that degrade the mucin glycoproteins. This enables them to adhere efficiently to the host and thereby cause infection within the host. The glycoside hydrolases produced by bacteria, cleave the carbohydrate chains in mucin and hydrolyse them to produce monosaccharide units. These monosaccharides are then used by the organism as a source of energy and nutrition (Roberton & Stanley, 1982).

Mucin, being the main component of mucosal surfaces acts as a point of adherence for the bacteria where they grow upon the glycoside residues of the complex glycoprotein oligomer by producing glycoside hydrolases and this leads to further damage of the mucosal surface and impairment of the membrane function (Salyers *et al., 1977*). *B. fragilis* has also been known to ferment mucin to release sulphates which favour the growth of other sulphate reducing bacteria. They break down sulphates into sulphides which can be toxic to the human system and this suggests that *B. fragilis* may possess sulfatase activity (Willis *et al.,* 1996).

Mucin, a high molecular weight glycoprotein that is heavily glycosylated and phosphorylated acts as a potential target for commensal bacteria like sulphate reducing bacteria which metabolise the sulphates and phosphates to produce sulphide ions and toxic hydrogen sulphide. This alters the environment in the gut and encourages the growth of other intestinal infection causing bacteria. According to previous studies, the sulphation of the gastrointestinal tract increases from 1-3% in the stomach to 3-5% in the colon (Willis *et al.*, 1996). The intestinal microflora depend upon the type of nutrition and it has been observed that changes in nutrition can affect metabolism and cause an imbalance or change in the microflora of the gut. The other factor that contributes to the type of microflora present in the gut is the oxygen concentration. The upper part of the gastrointestinal tract is colonised by aerobic flora whereas the lower part of the gut is colonised by anaerobes. This may be due to the decrease in concentration of oxygen from the duodenum to the colon (Serino *et al.*, 2009). Evasion of the host immune system results in the development of an infection. Previous studies have shown that micro-organisms have the ability to colonise the reproductive tract by producing enzymes like mucinases and sialidases which degrade the protective mucin layers in the cervix (Wiggins *et al.*, 2001).

Some of the reasons for the colonisation of mucosal surfaces by bacteria may be due to chemotaxis and motility. Previous studies have shown that mutants lacking motility and chemotactic abilities exhibited reduced virulence (Laux *et al*., 2005). Bacterial flagella can function as adhesins too. Adherence of bacteria to the mucous layer receptors prevents their adherence to the cell receptors of the intestine. But the process behind how this adherence enhances colonisation is not known. Hence the cell surface structures that are expressed in bacteria during colonisation play an important part in understanding the mechanism of pathogenesis (Laux *et al*., 2005). The *B. fragilis* genes BF0855 and BF3763 which were cloned and expressed in *E. coli* encoded the putative expression of the enzymes α-fucosidase and α-glucosidase respectively. These enzymes have the ability to hydrolyse oligosaccharides into simple sugars like glucose or galactose and may play an important role in the degradation of complex polysaccharides like mucins. These enzymes may be associated with the differential expression of proteins in mucin enriched media and was one of the main reasons for their purification and characterisation in this study.

The study of the genome of organisms is termed as 'genomics' and the sequencing of genomes forms the basic blue print behind the study of proteins. There may be several modifications, interactions and other complex structures formed by proteins but all of these may be linked to the genome through various factors like the environment (Zivy & de Vienne, 2000).

4.6 Proteomics

Proteomics is a field of science that merges the study of genes and proteins employing technologies such as two dimensional gel electrophoresis and mass spectrometry. It is basically a genomic study of the functions and expression of proteins where the two dimensional gel electrophoresis is used to measure the expression of a protein and mass spectrometry is used in protein identification and characterisation. The advancements in the field of genomics has led to the more frequent employment of proteomics to study proteins/enzymes, and has led to advancements in antibacterial drug discovery, identification of drug targets and the

mechanisms of resistance exhibited by bacteria and various other characteristics related to it (Brotz-Oesterhelt *et al*., 2005).

The term ‘proteome’ has been derived from the terms ‘protein’ and ‘genome’. There are several factors that can influence the proteome of an organism, namely stress, culture conditions, environment, metabolism, temperature, drugs, illness and so on. The genome encodes the cell specific expression of proteins based on the effect of these factors. Protein modifications also act as important factors that contribute to the functions of a cell and the study of all the proteins produced by a particular species in an environment is collectively termed as proteomics. The five major aspects that contribute to research in proteomics include mass spectrometry based proteomics, proteome-wide biochemical assays, systematic structural biology and imaging techniques, proteome informatics and clinical applications of proteomics (Tyers & Mann, 2003). Proteomics serves as a tool in studying post translational modifications that alter physical and chemical properties of proteins including folding, conformation, stability and functions. Proteomic techniques like stable isotope labelling and novel mass spectrometric peptide sequencing can be used to characterise modifications in proteins like phosphorylation or in membrane proteins. Proteomics has also been used as a valuable tool in the identification of protein isoforms in eukaryotes (Blakeley *et al*., 2010). Alternate splicing of mRNAs result in a major difference between the sequences of genes present and the translated protein isoforms that can be of varying complexities. Protease associated degradation of proteins acts as an important factor in regulating signalling pathways and other physiological processes where selective affinity purification and tandem mass spectrometric analysis of resulting peptides can be used to study the processing events that may be taking place during cell signalling (Xu *et al*., 2009).

4.7 Transcriptomics

Analysing or studying proteins that are expressed in a cell through the mRNA present in the cell is called transcriptomics. Transcriptomics forms the link between proteomics and genomics since messenger RNA is the link between genes and proteins (Hegde *et al*., 2003). Transcriptomics works based on the principle that genes showing similarities in their expression patterns may be functionally related too and hence it may be of importance to study the genes that control these expression profiles. Mutations in gene sequences can be used to study their potential as prospective drug targets, determining the functions of DNA sequences with no known activity based on their similarity to other conserved genes that may be homologous in nature (Twyman, 2003).

4.8 Two dimensional polyacrylamide gel electrophoresis (2D-PAGE)

This technique plays as a major role in the separation of mixtures of proteins. It is mainly composed of two techniques namely isoelectric focussing (IEF) and sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE).

IEF is the first dimensional separation of the protein based on the net charge that it possesses. A high voltage is applied to the proteins of interest which are to be separated and an immobilised pH gradient technique is used in the separation of proteins using specific reagents. Proteins usually possess a charge and tend to migrate towards the oppositely charged electrode in the presence of an electric field. When a positively charged protein migrates towards the cathode, it passes through an increasing pH gradient which causes an overall decrease in the charge of the protein. Eventually, the protein reaches a pH region that corresponds to its isoelectric point where the net charge becomes zero and migration stops. This results in the formation of a protein band at that particular point. Following the separation based on charges, they are then run on a second dimension SDS-PAGE where the separation is based on the molecular weight and size. This technique employs the use of a denaturing anionic detergent, sodium dodecyl sulphate, in the effective separation of proteins (Gorg *et al*., 2004). The anionic detergent confers a negative charge to the protein in proportion to its length by wrapping itself around the

polypeptide backbone. When an electric field is applied, the proteins migrate towards the positive electrode. The size of the pores in the polyacrylamide gel are user-defined and differentially retard the migration of proteins based on their size. Smaller protein molecules tend to travel farther down the gel when compared to larger protein molecules that remain closer to the point of origin resulting in effective separation of proteins. Therefore, by combining the separation of proteins based on their charge, or isoelectric point, and size, or relative molecular mass, differential expression studies can be performed.

Previous proteomic studies have been done in analysing the oxidative stress response in *B. fragilis* since they are known to be strict anaerobes and it was found that these bacteria are capable of inducing OxyR regulon genes that respond to the presence of oxygen or hydrogen peroxide in the environment and this might play an important role in their ability to survive within the tissues in the colon (Rocha *et al.*, 2003).

4.9 Mass spectrometry

Separation of molecules on the basis of their mass to charge ratio following ionisation is called mass spectrometry where a mass spectrum is generated based on the intensity of ionisation. The detected spectrum is then compared to a standard database to determine the identity of the complex peptide mixture. The sample is first passed through the ioniser that charges the molecules and the ions that are generated, enter the analyser where the ions are separated in vacuum. The individual ions from the analyser are then passed on to the detector where a spectrum based on the intensity of the ions is generated. The height of the peak in a spectrum is referred to as intensity. Mass spectrometry acts as a major tool in genomics and proteomics where complex molecules and protein mixtures can be identified and analysed (Liebler, 2002). Identification of proteins from a database is called peptide mass fingerprinting (PMF) or peptide mass mapping. Some methods of analysis also involve the comparison of a theoretical calculated spectrum of a database to an experimental spectrum generated from the mass spectrometer. One of the common methods used for the identification of proteins is the tandem mass spectrometry which consists of two phases. The first phase constitutes the ionisation

of peptides to produce a spectrum that is used for peptide mass fingerprinting. Some of the peaks from this spectrum are selected based on the intensity and subjected to the second phase of dissociation. The ions are transmitted through a high pressure region of the tandem mass spectrometer containing gas molecules. The gas molecules collide with the ions and further fragment them into ions of varying masses called collision induced dissociation (CID) (Liptak, 2005). Since fragmentation can occur between any two residues along the peptide, differing masses of ions are generated and can be termed as b and y ions (Refer to Figs. 7 and 8). These ions are then mapped to their respective proteins using a database like MASCOT (Johnson, 1987).

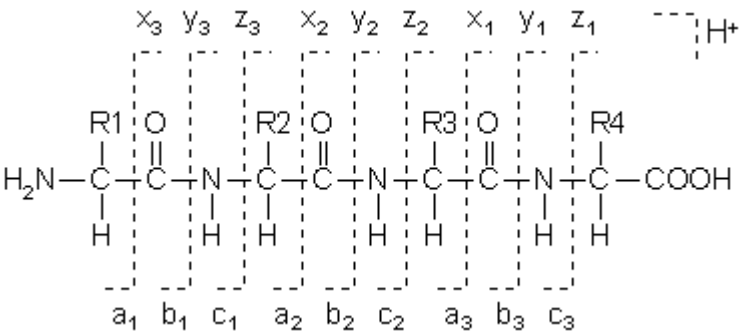


Figure 7 represents the fragment ions in an MS/MS spectrum (Anon, 2010)

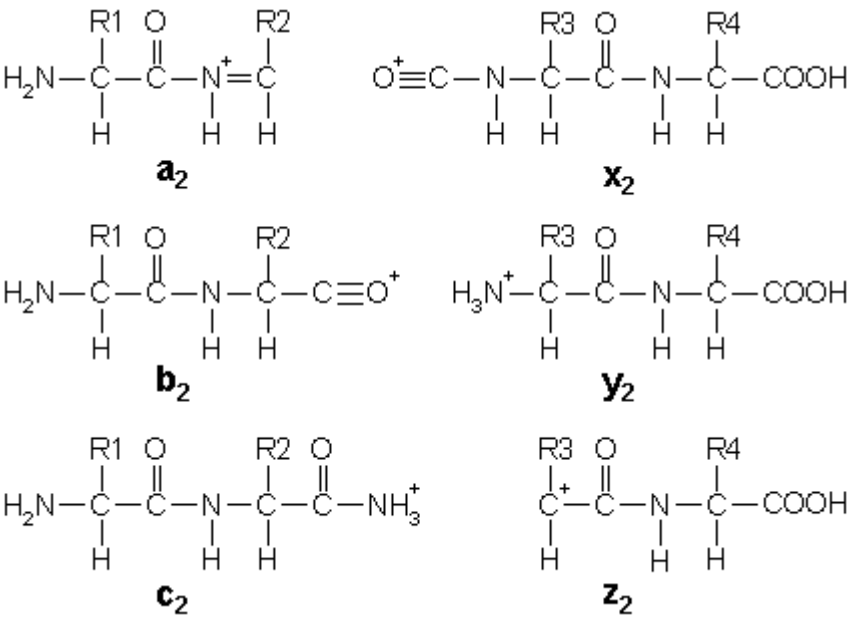


Figure 8 shows the structures of six singly charged sequence ions (Anon, 2010).

4.10 Glycoside hydrolases

Glycoside hydrolases or glycosidases are enzymes that catalyse the hydrolysis of sugars and other complex polysaccharides into smaller and simpler sugar units by hydrolysing their glycosidic linkages. These simple sugars or end products help in carbohydrate metabolism and are used as a source of carbon and energy in bacteria and higher animals. Glycosidases in association with glycosyltransferases, form the main machinery for the synthesis and breakdown of glycosidic bonds. They are produced as extra and intracellular enzymes in bacteria that help to acquire nutrients by breaking down complex compounds into simple sugars. Within the intestinal tract, they are capable of degrading carbohydrates like lactose, starch, sucrose and trehalose.

Glycoside hydrolases can be classified based on the stereochemical outcome of their hydrolysis reactions as inverting or retaining mechanisms. They act on oligo and polysaccharide chains by hydrolysing the glycosidic linkages in the non-reducing end or the middle of the chain. Hence this can be used to classify them as exo-acting or endo-acting glycoside hydrolases. The exo-acting enzymes are capable of cleaving residues from the ends of the polysaccharide chains whereas endo-acting enzymes remove residues from the middle or internal positions in a random manner resulting in several oligosaccharides of different sizes. They have also been classified on sequence based methods where sequence similarity acts as a major factor. This series of sequence based classification of enzymes allows the prediction of their mechanism of hydrolysis, active site residues and possible substrates that can confer this activity. The CAZy website enzyme database is regularly updated and classifies enzymes based on their sequence and three dimensional structural similarities. The glycosidase sequence analysis and three dimensional structure comparison has led to the generation of a hierarchial classification of these glycoside hydrolases (http://www.cazy.org/fam/acc_GH.html).

4.10.1 Inverting mechanism in glycoside hydrolases

Reactions involving the inverting mechanism of hydrolysis of carbohydrates employ the use of two enzymatic residues where one acts as an acid and the other acts as a base. The reaction occurs in the presence of carboxylic acid at the active site. This type of catalysis can be termed as a general acid base assisted catalysis method (Sinnott, 1990).

Figure 9: Figure showing the catalytic mechanism in glycoside hydrolases (http://www.cazy.org/fam/ghf_INV_RET.html)

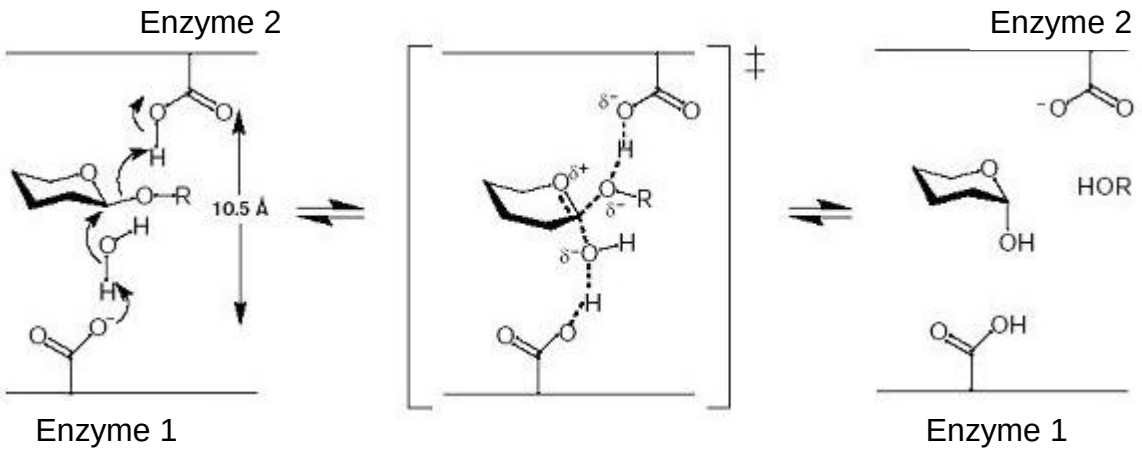


Figure 9.1: Figure showing the inverting mechanism of hydrolysis. Enzyme 1 acts as a base and the enzyme 2 acts as an acid by donating protons.

4.10.2 Retaining mechanism in glycoside hydrolases

The retaining mechanism involves a 2 step reaction and each step results in an inversion of compounds. There are two enzyme carboxylic acid residues present where one acts as a nucleophile and the other acts as an acid or base by donating or accepting protons respectively. When the nucleophile attacks the anomeric centre in the first step, a glycosyl intermediate is formed. The acidic carboxylate acts as a proton donor and assists the reaction. The second step involves the hydrolysis of the

glycosyl enzyme intermediate by the deprotonated acid carboxylate which acts as a base to produce the hydrolysed product. The reaction also involves the use of a nucleophilic water molecule. The following figure represents the inverting and retaining mechanism of glycoside hydrolase enzymes.

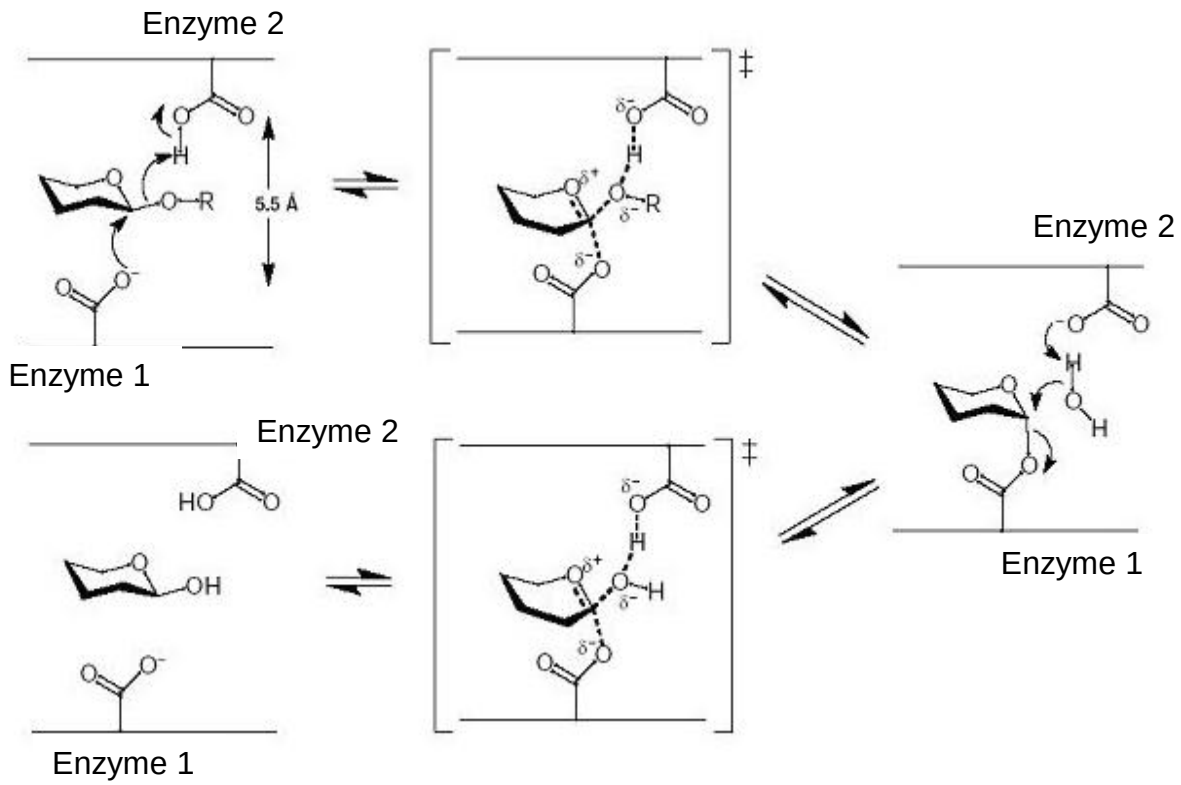


Figure 9.2: Figure showing the retaining mechanism of hydrolysis. The enzyme 1 acts as a nucleophile and the enzyme two acts as an acid or a base. A glycosyl enzyme intermediate is formed before the substrate is converted into the hydrolysed product.

4.10.3 Carbohydrate active enzymes (CAZy) based classification of enzymes

The CAZY website classifies enzymes into five main categories namely glycoside hydrolases (GHs), polysaccharide lyases (PLs), glycosyltransferases (GTs) and carbohydrate esterases (CEs) and carbohydrate binding modules (CBMs). Substrate specificity and molecular mechanisms contribute to some of the main factors considered for classification. The relationship between sequences and their folding similarities led to the consideration of amino acid sequence similarities as a major

factor. The active sites remain highly conserved and are maintained with integrity even though the structures of enzymes may differ due to evolutionary divergence. The families have also been grouped into ‘clans’ based on the strong conservation of the folds of proteins even though sequences were not found to be as conserved as the folds of proteins. This grouping of families into clans would help to overcome structural and sequential resemblances or relatedness of enzymes to more than one family. The other factor that supported the grouping of families into clans was the improved sensitivity of sequence comparison methods. As of May 2009, 115 families of enzymes have been identified and classified in the CAZY website (http://www.cazy.org/fam/acc_GH.html).

4.10.4 Active site topology

Glycoside hydrolases have the presence of specific active sites based on the position of hydrolysis of the polypeptide chain. These may be of three types namely, the pocket site which is characteristic of exo enzymes, the cleft structure that occurs in endo enzyme hydrolysis and the tunnel topology in exo enzyme hydrolysis (refer to the Fig. 10). While the pocket topology binds to the ends of the polysaccharide chains, the cleft topology binds chains randomly within a cleft. The tunnel topology being very similar to the cleft topology has additional polypeptide loops which are present on the top of the cleft giving it a tunnel like appearance. The end of the polypeptide chain enters the tunnel and the digested peptides are released out at the other end. In the case of the cleft active site, random hydrolysis of the polypeptide chain results in the release of oligosaccharides or various lengths (Davies & Henrissat, 1995).

Figure 10: Figure showing the active site topology in glycoside hydrolases (Davies & Henrissat, 1995)

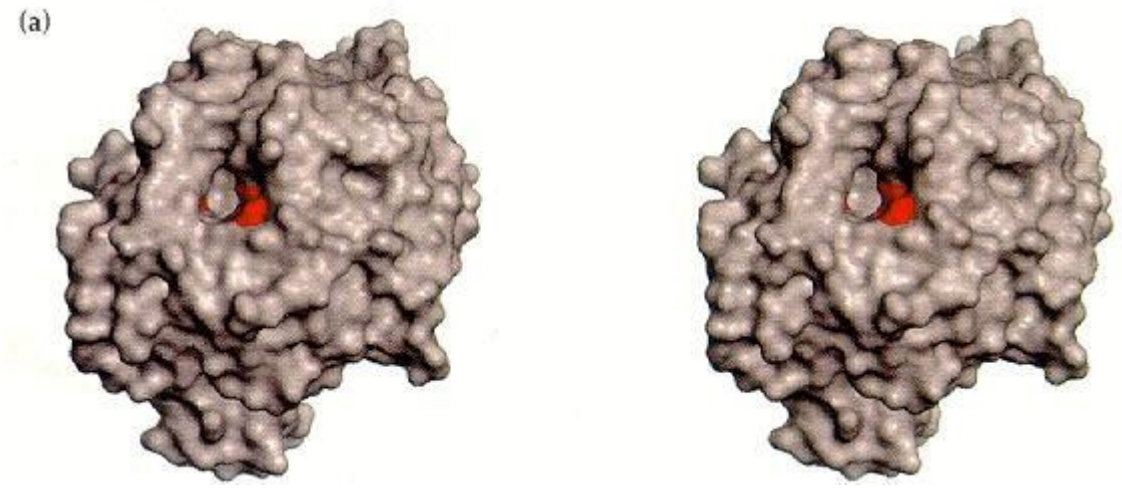


Figure 10.1: Pocket topology. Coloured area indicates the substrate binding site of the enzyme.

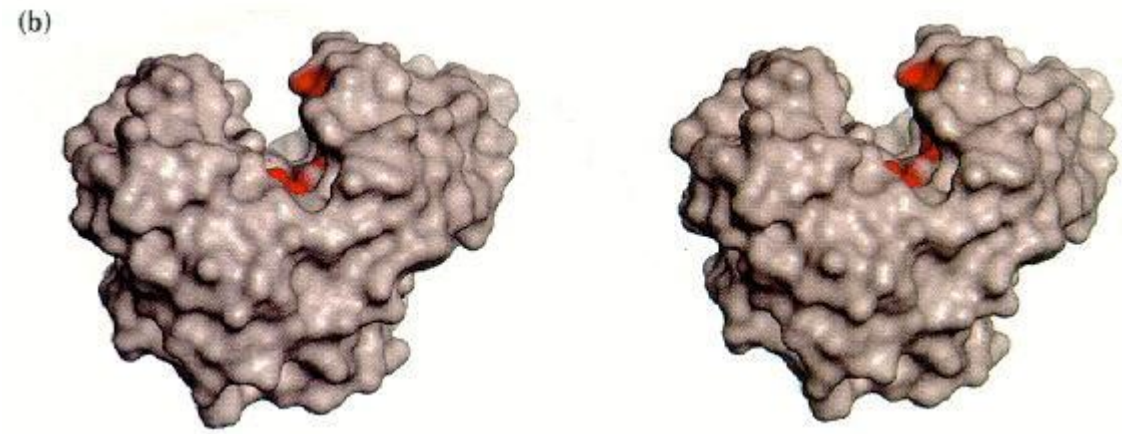


Figure 10.2: Cleft topology. Coloured area indicates the substrate binding site of the enzyme.

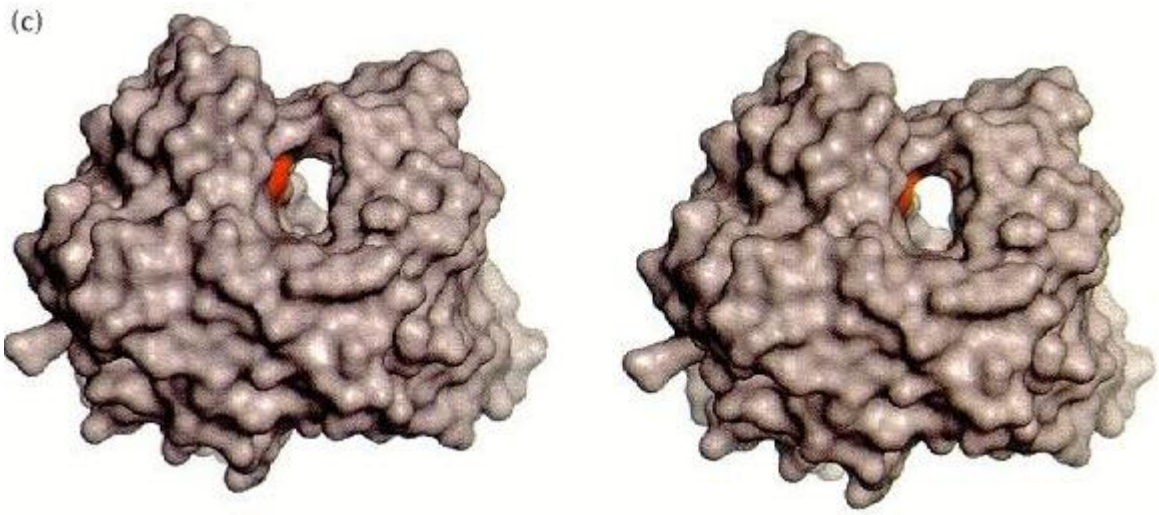


Figure 10.3: Tunnel topology. Coloured area indicates the substrate binding site of the enzyme.

4.10.5 Genomes of *B. fragilis* and *E. cancerogenus*

Completed published sequences of the genome of a number of different bacteria are readily available online. As of 11th May, 2009, 992 genomes have been published and are available via the Gold genome database. The online database reports 2523 ongoing bacterial genome sequencing projects along with 1029 eukaryotic genomes and 96 archael ongoing projects. The total number of ongoing genome projects has been reported to be 4807 when compared to 2006 when it was 2172 (http://www.genomesonline.org/gold.cgi).

The sequencing of genomes provides important information regarding the function of proteins. Once a genome is sequenced, the DNA is annotated as putative open reading frames (ORFs) containing information about the coded regions including exons, introns, promoters and so on (Zhang, 2006). Based on the translation of these ORFs, proteins or amino acid sequences can be obtained and searched against a sequence database to determine a list of proteins with the highest scores. Protein hits that show a very high degree of similarity or homology indicate a functional similarity between the two proteins. This can also be used to study the similarities and relationships between different bacterial species. Some of the problems that may be associated with this approach are that one particular gene

could encode for more than one protein and this could create confusion during analysis (Pennington & Dunn, 2001).

The genome of *B. fragilis* NCTC 9343 has been sequenced and published by the Sanger Institute (http://www.sanger.ac.uk/Projects/B_fragilis/). The genome of *E. cancerogenus* ATCC 35316 is still being sequenced and is available through the NCBI website but work is still ongoing. The sequencing information is available from the Washington University or through the Gold genomes online database (http://www.genomesonline.org/gold.cgi?want=Bacterial+Ongoing+Genomes).

Genomes of various species of *Bacteroides* namely *B. vulgatus*, *B. diastonis* were compared with *B. fragilis* and *B. thetaiotomicron* in a study. The conserved gene sets were grouped into Clusters of orthologous groups. Gut associated Bacteroidetes proteomes showed the predominant expression of conserved genes in carbohydrate transport and metabolism followed by cell wall or membrane or envelope biogenesis. The shared proteome of Bacteroidetes showed the expression of genes involved in amino acid transport and metabolism followed by translation, ribosomal structure and biogenesis genes (Xu *et al.*, 2007).

4.10.6 Glycoside hydrolase families GH95 and 97 and the putative activity of their enzymes

The GH95 BF0855 gene encodes for a conserved hypothetical protein containing 755 amino acids with a molecular weight of 86 kDa. The CAH06598.1 protein exhibits putative α1-2, L-fucosidase and α-L-fucosidase actvitiy and the catalytic nucleophile is asparagine that is activated by aspartic acid residues (Cerdeno-Tarraga *et al.*, 2005). The catalytic proton donor has been identified to be glutamic acid residues. Previous structural studies of these enzymes in *Bifidobacterium bifidum* have shown that the enzyme has a high degree of similarity to the immunoglobulin Ig domains with similar folding patterns. They possess α/α six barrel or catalytic TIM type barrel domain and are active against substrates like 2'-fucosyllactose and lacto-N-fucopentaose I (Katayama, 2004) .

The GH97 BF3763 gene encodes for a putative export protein containing 649 amino acids with a molecular weight of 74 kDa. The CAH09443.1 protein is known to be a

putative α-glucosidase or α-galactosidase. Enzymes that exhibit inverting mechanisms of catalysis contain glutamic acid as nucleophiles and those that exhibit retaining mechanisms of catalysis possess aspartic acid residues as nucleophiles (Gloster *et al.*, 2008). Previous studies of the starch utilisation system in *Bacteroides thetaiotomicron* showed an inversion mechanism of catalysis liberating a β-anomer of glucose (Kitamura *et al.*, 2008). The possess α/β eight barrel or catalytic TIM barrel type domains and the amino acids of the active site are located at the C-termini of the β-strands. These proteins exhibit a high degree of similarity to proteins of other families that exhibit a retaining mechanism of glycoside bond hydrolysis (Naumoff, 2005).

The BF0855 gene from *B. fragilis* NCTC 9343 which was cloned and expressed in *E. coli* with a putative activity of α-L-fucosidase or 1, 2 α-L- fucosidase catalyses the reaction between α-L- fucoside and water resulting in the formation of L- fucose as an end product. These enzymes belong to the family of glycoside hydrolases that hydrolyse O and S glycosyl compounds and are also known as α-L- fucoside fucohydrolase. It is involved in n- glycan and glycan structure degradation and may be involved in the degradation of mucin by bacteria. 1, 2-α-L- fucosidase catalyses the following reaction:

Methyl- 2- para- L- fucopyranosyl- β- D- galactoside + H_2O ↔ L- fucose + methyl β-D- galactoside.

5 Aims of the study

This study aims at understanding the differences in protein expression in *E. cancerogenus* ATCC 35316 and *B. fragilis* strain NCTC9343 in mucin-supplemented media in comparison to basic semi-defined media containing glucose as the carbohydrate source. These are commensal organisms that act as opportunistic pathogens and can cause infections within the host when the immune system is impaired.

Two dimensional gel electrophoresis technologies would be employed, where the proteins were first separated in the first dimension based on their charge and pH using isoelectric focussing and then separated using the second dimension polyacrylamide gel electrophoresis based on their molecular weight. Protein expression profiles would be generated for both bacteria grown in the presence and absence of mucin and compared. Comparative analysis of gels would be used to determine the protein spots that show a differential expression in the presence of mucin.

The protein spots of interest will be excised and subjected to in-gel trypsin digestion where the peptides produced from the degradation of proteins would be analysed using a nano flow liquid chromatography coupled mass spectrometer. This would help to identify the protein based on the mass spectrum (m/z ratio) generated from the MS-MS fragmentation of peptide ions. The identification of the differentially expressed proteins may provide an insight into the indirect association of bacteria and their colonisation of mucin. It may also be interesting to use mucin-coated cell culture plates to study the interaction between bacteria and mucin in the presence of a host immune response.

This study also aims to express, purify and crystallise the proteins CAH06598 and CAH09443 from the glycoside hydrolase families 97 and 95 in *B. fragilis* strain NCTC 9343 which produce the enzymes with the putative functions of α-fucosidase and α-glucosidase respectively. The recombinant enzyme activity was analysed against different chromophoric and fluorogenic substrates. Crystallisation of proteins would be useful in determining its three dimensional structure and obtain a detailed understanding of its active sites.

6. Materials for the Proteomics experiment

6.1 Culture Conditions

Bacteroides fragilis strain NCTC 9343 and *Enterobacter cancerogenus* ATCC 35316 were cultured in Columbia agar plates (CBA). A single colony from the plate was used to inoculate 20 mL of sterile anaerobic broth media (Bac T/ ALERT SN). The medium was incubated overnight in an anaerobic chamber at 37 °C. A loopful of the liquid culture was used to streak a CBA plate (Figure 3). The purity of the subculture was confirmed by Gram staining which showed Gram-negative rods.

6.2 Anaerobic chamber

The bacteria was cultured and maintained in an anaerobic chamber that was flushed with 80% nitrogen gas, 10% carbon dioxide and 10% hydrogen (Cox & Mangels, 1976).

6.3 Minimal media used for the growth of *B. fragilis*

Minimal media used for the growth of *B. fragilis* according to the method of Varel and Bryant (1974)

Per 100 ml	
Glucose	0.5 g
Mineral solution[1]	5.0 mL
Hemin solution[2]	0.1 µL
Resazurin solution[3]	0.1 µL
Volatile fatty acid solution[4]	0.45 mL
Vitamin B solution[5]	0.5 mL
$FeSO_4$-$7H_2O$	0.4 mg
6 mM $(NH_4)_2SO_4$	5.0 mL
Casitone (Difco)	0.2 g
Amino acid mixture[6]	20 mL
2.5% (w/v) Cysteine-HCl solution	2.0 g
8% (w/v) Na_2CO_3 solution	5.0 mL

The medium was made up with distilled water and autoclaved before use. The pH of the medium was adjusted to 7.0 and all components were added before autoclaving except sodium carbonate and cysteine. The medium was incubated overnight in the anaerobic chamber at 37°C in the absence of oxygen before inoculation of starter cultures.

6.3.1 Stock solutions

6.3.1.1 Mineral solution[1]

Per litre

KH_2PO_4	18 g
NaCl	18 g
$CaCl_2$-$2H_2O$	0.53 g
$MgCl_2$-$6H_2O$	0.4 g
$MnCl_2$-$4H_2O$	0.2 g
$CoCl_2$-$6H_2O$	0.02 g

6.3.1.2 Hemin stock solution[2]

A stock concentration of 0.1% (w/v) of hemin solution was obtained by dissolving 0.1 g of hemin in 1 mL of 0.1 M sodium hydroxide and then diluted to 100 mL with distilled water and appropriate dilutions of the stock were added to obtain a final concentration of 0.0001% (w/v) of hemin solution.

6.3.1.3 Resazurin solution[3]

A stock concentration of 0.1% (w/v) of resazurin solution was prepared by dissolving 0.1 g of resazurin in 100 mL of sterile 18.2 MΩ/cm and appropriate dilutions of the stock were added to obtain a final concentration of 0.001% (w/v) of resazurin solution.

6.3.1.4 Volatile fatty acid solution[4]

Per 100 mL

Acetic acid	36 mL
Isobutyric acid	1.8 mL

n-Valeric acid	2.0 mL
DL-2 Methylbutyric acid	2.0 mL
Isovaleric acid	2.0 mL

6.3.1.5 Vitamin B solution[5]

Per 100 mL

Thiamin-hydrochloride	20 mg
Calcium-D-pantothenate	20 mg
Nicotinamide	20 mg
Riboflavin	20 mg
Pyridoxine-HCl	20 mg
p-aminobenzoic acid	1 mg
Biotin	0.25 mg
Folic acid	0.25 mg
Vitamin B12	0.1 mg

6.3.1.6 Amino acid mixture[6]

Per 50 mL

L-histidine-hydrochloride	25 mg
L-tryptophan	25 mg
Glycine	25 mg
L-tyrosine	25 mg
L-arginine-hydrochloride	50 mg
L-phenylalanine	50 mg
L-methionine	50 mg
L-threonine	50 mg
L-alanine	50 mg
L-lysine	75 mg
L-serine	75 mg
L-valine	75 mg
L-isoleucine	75 mg
L-proline	75 mg
L-aspartic acid	75 mg
L-leucine	100 mg

L-glutamic acid	224 mg

6.4 Semi defined growth media

Per 100 mL

Casitone	0.5 g
Yeast extract	0.5 g
Ammonium sulphate	0.08 g
Salts solution A*	0.04 mL
Salts solution B *	0.005 mL
Hemin solution	0.01 mL
Glucose	0.5 g
Sodium carbonate	0.8 g
Cysteine-HCl	0.05 g

The medium was adjusted to a pH of 7 and all components were added before sterilisation except glucose and Na_2SO_4. Sterile glucose and sodium carbonate solutions were prepared separately and added aseptically to the medium. The medium was incubated overnight at 37 °C in the anaerobic chamber before inoculation with the starter cultures. The hemin solution was prepared by dissolving 50 mg of hemin in 1 ml of 0.1 M NaOH which was diluted with distilled water to 100 mL. In 100 mL of the mucin supplemented semi-defined media, 0.5 g of mucin Type II or III were added to the media in addition to 0.5 g of glucose and autoclaved. The 0.5 g refers to the total amount weighed out and added rather than the concentration of sugars in mucin.

Sigma-Aldrich porcine gastric mucin Type II and Type III were used in the proteomics experiments. A quantity of 0.5 per of the glycoprotein was weighed out and added to 100 mL of the semi-defined medium and dissolved by vigorous shaking. Mucin was added to the semi-defined medium just before autoclaving and care was taken to avoid inhaling the chemical dust.

6.4.1 Stock solutions

6.4.1.1 Salts solution A

Per litre

$CaCl_2$	0.2 g
$MgSO_4$	0.2 g
K_2HPO_4	1.0 g
KH_2PO_4	1.0 g
NaCl	2.0 g
$NaHCO_3$	10.0 g

6.4.1.2 Salts solution B

Per litre

$FeSO_4$-$7H_2O$	2.0 g
CoCl	0.08 g

6.5 Gram staining reagents

Crystal violet	5 mL
Gram's iodine	5 mL
Acetone	5 mL
Safranin solution	5 mL

6.6 SDS-PAGE Loading buffer

Per 10 mL

60 mM Tris Base pH 6.8	0.6 mL
50% (w/v) glycerol	5 mL
10% (w/v) sodium dodecyl sulphate (SDS)	2 mL
14.4 mM β-mercaptoethanol	0.5 mL
1% (w/v) Bromophenol blue (BPB)	1 mL

Stored as 1 mL aliquots at -20 °C

6.7 Protein size standards

High molecular weight standard (M.W. 36, 45, 55, 66, 84, 97, 116 and 205 kDa)

Low molecular weight standard (M.W. 20, 24, 29, 36, 45 and 66 kDa)

To obtain a final concentration of about 2.0-3.5 mg/mL, the lyophilised standards were reconstituted in 100 µL of 18.2 MΩ/cm water and aliquoted out in 4 µL quantities into 1.5 mL microcentrifuge tubes and stored at -20°C. Refer to appendix D for the list of proteins used as a source for producing these size standards.

6.8 Buffers and solutions used for the Proteomics experiment

6.8.1 Cell resuspension buffer

The cell resuspension buffer has the same recipe as that of the phosphate buffered saline (6.8.2)

6.8.2 Phosphate buffered saline (PBS)

Per litre (pH 7.0)

NaCl	8 g
KCl	0.2 g
KH_2PO_4	0.2 g
Na_2HPO_4-$12H_2O$	1.44 g

6.8.3 Lysis solution

Per 10 mL

Urea	4.8 g
3-[(3-Cholamidopropyl) dimethylammonio]-2-hydroxy-1-propane sulphonate (CHAPS)	0.4 g

The solution was made up in a sterile universal and aliquoted out into microcentrifuge tubes in 1 mL quantities. This was then stored at -20°C and 2 µL of IPG buffer was added to every 100 µL before use.

6.8.4 Rehydration solution (for IPG strips pH 3-10 or 4-7)

Per 10 ml

Urea	4.8 g
CHAPS	0.4 g
1% (v/v) BPB	20 µL

1% (w/v) bromophenol blue solution was prepared by dissolving 1 g of bromophenol blue dye in 100 mL of 18.2 MΩ/cm water.

The rehydration solution was divided into 700 µL aliquots and stored at -20°C. Prior to use, 14 µL IPG buffer (pH 3-10 or 4-7) and 1.4 mg of ditiothreitol (DTT) was added to each tube to give a final concentration of 2% (v/v) and 0.2% (w/v) respectively.

Isoelectric focusing or the first dimension separation of proteins was carried out using a Multiphor II electrophoresis system involving the use of other components listed in the Appendix section B.

Main parts include the anode and cathode electrodes, tray and electrode holder, dry strip aligners, IEF electrode strips.

6.8.5 Equilibration solution

Stock solution:

Per 200 mL

1.5 M Tris-HCl (pH 8.8)	10 mL
Urea	72 g
Glycerol	69 mL
SDS	2 g
1% (v/v) BPB	200 µL

The stock solution was stored at -20°C in 20 mL aliquots.

6.8.5.1 Equilibration buffer with DTT

Per 20 mL

Equilibration stock solution	20 mL
DTT	0.2 g

6.8.5.2 Equilibration buffer with Iodoacetamide (IAA)

Per 20 mL

Equilibration stock solution	20 mL
Iodoacetamide (IAA)	0.9 g

6.9 Second dimension-SDS-PAGE

The SDS-PAGE was performed using a Protean II XL 2-D cells (Bio-Rad).

4 mm xi clamp notch vs 13 mm XL clamp notch

19 mm xi spacer vs. 8 mm XL spacer

181 mm xi core gasket vs. 198 mm XL core gasket

153 mm xi comb vs. 184 mm XL comb

The main parts include tank and lid, central cooling core, casting stand, sandwich clamps, alignment card and combs.

Small gels- glass plate sizes were 10.1 X 8.2 cm

Large gels- glass plate sizes were 20 X 20 and 20 X 22 cm separated by spacers of 1.5 mm

6.9.1 14% (w/v) Resolving gel components

18.2 MΩ/cm water	58 mL
1.5 M TRIS-HCl, pH 8.8	37.5 mL
10% (w/v) SDS stock	1.5 mL
40% (v/v) solution (37.5:1 acrylamide: bisacrylamide)	52.2 mL
10% (w/v) APS	750 µL
N, N,N',N'-Tetramethylethylenediamine (TEMED)	75 µL

6.9.2 4% (w/v) Stacking gel components

18.2 MΩ/cm water	12.85 mL
0.5 M TRIS-HCl, pH 6.8	5 mL
10% (w/v) SDS stock	200 µL
40% (v/v) solution (37.5:1 acrylamide: bisacrylamide)	1.95 mL
10% (w/v) ammonium persulphate (APS)	100 µL
TEMED	20 µL

1% (w/v) BPB	20 µL

6.9.3 Agarose sealing solution

Per 20 mL

Running buffer (1 X)	19 mL
Agarose	0.1 g
1% (w/v) BPB	40 µL

Protein size standards were the same as those mentioned in the above sections (Refer to Appendix D for details).

6.9.4 Staining and destaining solutions

6.9.4.1 Colloidal Coomassie blue stock

Per litre

Ammonium sulphate	100 g
Phosphoric acid	20 mL
Coomassie blue G 250	1 g

20 mL of 18.2 MΩ/cm was used to dissolve 1 g of Coomassie blue G 250. A 20 mL solution of phosphoric acid was used to dissolve 100 g of ammonium sulphate and 18.2 MΩ/cm water was added if necessary. The two solutions were then mixed together and made up to 1 litre with 18.2 MΩ/cm water.
The stock solution was stored at room temperature until use.

6.9.4.2 Fixing solution (Colloidal Coomassie blue staining)

Per litre

Methanol	500 mL
Glacial acetic acid	120 mL

6.9.4.3 Staining solution (Colloidal Coomassie blue staining)

The stock solution was shaken well to suspend the Coomassie blue stain and four parts of the Coomassie stock was diluted with one part methanol before use.

6.9.4.4 Fixing solution (Fluorescent staining)

Per litre

Ethanol	400 mL
Acetic acid	100 mL

6.9.4.5 Staining solution (Fluorescent staining)

A 10 X concentration of the Bio-Rad Flamingo™ fluorescent gel stain was diluted to a 1 X working concentration of the solution with 18.2 MΩ/cm.

6.9.5 Visualisation of the stained gels

The Bio Rad Chemi doc XRS (Quantity One ™ software, Resolution: 1392 X 1040 pixels) was used to capture images of the gel using UV transillumination for fluorescence stained gels and the GS-800 densitometer (Resolution: 1360 X 1024 pixels) was used to scan Coomassie blue stained gels. Hard copies of the gel were printed out using the Mitsubishi Video Copy Processor attached to it.

Image acquisition for Coomassie blue stained gels was carried out using Bio-Rad GS-800 image densitometer and further analysis was done using the PDQuest™ Advanced v 8.0 software. Hard copies of the gel were produced using the Mitsubishi Video Copy Processor (K65HM-$_{CE}$ / High density type, 110 cm X 21 m). The raw image data which was obtained from scanning the gels using the GS-800 densitometer (Bio-Rad Quantity One ™ software) was uploaded into the other two softwares, Ludesi Redfin and Non-linear Dynamics Progenesis SameSpots and used in analysis.

6.10 Digestion of proteins for Mass Spectrometric analysis

6.10.1 Protein digestion in solution

6.10.1.1 Dissolve solution

Per 100 mL

SDS	0.1 g
Tris base	6.06 g
Dithiothreitol (DTT)	77.1 mg

The solution was adjusted to a pH of 8 using HCl and stored at room temperature.

6.10.1.2 Trypsin stock (1 µg/µL) (Promega and NEB)

Vial containing the lyophilised powder was stored at -20°C

Trypsin lyophilised powder 20 µg/mL	100 µg
Glacial acetic acid (50 mM)	100 µL

The trypsin powder was dissolved in 50 mM glacial acetic acid and stored at -20°C for upto 1 month or at -80°C for a long term.

6.11 In-gel protein digestion reagents

6.11.1 100 mM NH_4HCO_3

Per 10 mL

NH_4HCO_3	79 mg

6.11.2 50 mM NH_4HCO_3

Per 10 mL

100 mM NH_4HCO_3	5 mL

6.11.3 Trypsin solution (20 µg/mL)

Per 100 µL

1 µg /µL trypsin stock	2 µL
50 mM NH_4HCO_3	98 µL

6.12 Buffers used in LC-MS analysis

6.12.1 Start buffer A

Per 100 mL

LC-MS Grade Water	95 mL
LC-MS Grade Acetonitrile	5 mL
Formic acid	0.1 mL

6.12.2 Start buffer B

Per 100 mL

LC-MS Grade Acetonitrile	95 mL
LC-MS Grade Water	5 mL
Formic acid	0.1 mL

The solution was made up fresh before use.

7 Methodology used for the proteomics experiment

7.1 Growth study in *B. fragilis*

Ampules of *B. fragilis* were ordered from the National Collection of Type Cultures (NCTC) and used to inoculate the anaerobic basal broth media and incubated in the anaerobic chamber at 37°C overnight. The bacteria were sub cultured in Columbia blood agar media or brain heart infusion broth media.

The initial growth curve experiment was carried out by growing the bacteria in minimal medium whose components have been mentioned in the Materials section 6.3. But since the growth in the medium was not found to be reasonable enough for analysis, a semi defined medium was used. The components of the media varied with respect to yeast extract being used as the source of vitamin B. The medium was autoclaved before use and incubated overnight in the anaerobic chamber before inoculation. Mucin and glucose were weighed out and added at 0.5 g per 100 mL to the semi-defined growth media.

The semi defined media supported quicker growth of bacteria. Experiments were performed by using different carbon sources.

The culture was incubated at 37°C in the anaerobic chamber. The anaerobic condition within the chamber was monitored by using resazurin indicator strips that turned pink in the presence of oxygen in the chamber environment. Refer to Appendix E1 for details.

7.2 Growth curve studies

Small volume (20 mL) overnight starter cultures were used to inoculate large volumes (450 mL) of the media and the optical density of the media were monitored at various time intervals by measuring the absorbance at 600 nm and plating out serial dilutions of the culture. The time periods that were used to monitor the growth in *E. cancerogenus* were 0 h, 2 h, 3 h, 4 h, 5h, 6 h, 7 h, and 8 h whereas the time periods in *B. fragilis* were 0 h, 4 h, 6 h, 21 h, 22h, 23 h, 24 h, 25 h, 26 h, 27 h, 28 h, 29 h, 30 h, 31h, 46 h, 47h.The growth of *B. fragilis* in the medium was found to be slower and hence more time points were monitored. For each and every time point, the culture was mixed well and 1 mL was transferred into a plastic cuvette with a sterile pipette tip. The OD was measured at 600 nm using a Cecil Spectrophotometer. 100 µL of the culture was spread out onto a sterile CBA agar

plate and incubated overnight at 37°C. 1 mL of the culture was used to perform appropriate serial dilutions into test tubes containing 9 mL of sterile distilled water. Care was taken to maintain aseptic conditions within the chamber while performing the experiment. All experiments were performed in triplicates and the average values were calculated. Growth curves were plotted out to determine the phase of growth of the bacterium.

7.3 Two dimensional polyacrylamide gel electrophoresis (2D-PAGE) methodology

Two dimensional gel electrophoresis comprised of a series of steps from the growth of bacteria, protein extraction and one dimensional iso-electric focussing to the two dimensional separation of proteins using SDS-PAGE. The methodology has been explained in detail in the following sections.

7.4 Growth of *E. cancerogenus* and *B. fragilis* in semi defined media

A loopful of the glycerol stock was used to streak out a plate of Columbia blood agar medium and incubated overnight at 37°C in an anaerobic chamber.

A single colony of the bacteria from the culture plate was used to inoculate 450 mL of the semi-defined medium supplemented with and without mucin the following day. The semi defined medium was incubated overnight in the anaerobic chamber before inoculation. The culture was incubated in the anaerobic chamber till the optical density reached a value of 0.7 at 600 nm.

7.5 Protein extraction

A volume of 100 mL of the culture was then centrifuged at 4000 x g for 30 min at 4°C. Following this, the pellets were resuspended in 5 mL of PBS and washed at 4000 x g for 10 min at 4°C. The washing step was repeated 3 times and the final resuspension step was performed in 1 mL of PBS buffer. To 2 µL of IPG buffer pH 4-7 or 3-10, 100 µL of lysis solution stock was added to prepare the lysis solution and 100 µL of this was added to each of the samples. The pellets were then ultra-sonicated on ice for 1 min at 10 s intervals and then centrifuged at 14000 x g for 30 min at 4°C. The pellets were then discarded and the supernatant was transferred into clean 1.5 mL microcentrifuge tubes.

The cell free extract was subjected to removal of contaminants using the 2D clean up kit. It also helped to get rid of residual polysaccharides. To 100 µL of the cell free extract, 300 µL of precipitant solution was added and incubated on ice for 15 min.

To this, 300 µL of the co-precipitant was added and vortexed well before centrifuging it for 5 min at 14000 x g. The supernatant was discarded and 40 µL of the co-precipitant was added to the pellets and incubated on ice for 5 min. This was then centrifuged at 14000 x g for 5 min and the supernatant was discarded. 25 µL of 18.2 MΩ/cm water was added and vortexed well. To this, 1 mL of chilled wash buffer and 5 µL of wash additive were added before incubating the tubes at -20°C for atleast 30 min. The tubes containing the proteins could be stored at -20°C for a maximum of one week with minimal protein loss. The solution was centrifuged at 14000 x g for 5 min after which the supernatant was removed and the pellets were allowed to dry for half a min.

The pellets were re-suspended in 350 µL of rehydration solution containing 0.001 g of DTT, 4 µL of BPB, 7 µL of IPG buffer and 350 µL of the rehydration stock solution. The solution was vortexed and centrifuged at 14000 x g for half a min to remove the insoluble residue and the solution was transferred into the wells of the reswelling tray.

IPG strips of 11 cm length, 3.3 mm width and 0.5 mm thickness (pH 3-10 or 4-7) which were stored in the freezer at -20°C were taken and the plastic was removed from its surface. The strips were layered in the wells of the reswelling tray in such a way that the gel end of the strip was facing the rehydration solution and the positive end of the strip was located at the lower end of the reswelling tray. Once the strips were layered in the wells, the surfaces of the strips were covered with Dry strip cover fluid and left to rehydrate overnight at room temperature.

7.6 Iso-electric focussing

The temperature was maintained at 20°C

pH intervals	Step	Voltage (V)	Time (h)	kVh	Current (mA)	Power (W)
4-7	1	500	0:01	-	2	5
	2#	500	6:00	3.0	2	5
	3	3500	1:30	3.0	2	5
	4	3500	8:00	30	2	5
		Total:	15:31			

This step was included only when the Multiphor II Electrophoresis system was left running overnight since this is an optional step.

Table 1: Table showing the parameters used in isoelectric focussing, pH 4-7

pH intervals	Step	Voltage (V)	Time (h)	kVh	Current (mA)	Power (W)
3-10	1	500	0:01	-	2	5
	2	3500	1:30	3.0	2	5
	3	3500	6:00	22.0	2	5
		Total:	7:31			

Table 2: Table showing the parameters used in isoelectric focussing, pH 3-10

The Multiphor II from Amersham Biosciences was used to run the 1st dimension isoelectric focussing.

Two electrode contact strips of 11 cm were cut out and moistened with 1 mL of 18.2 MΩ/cm water. The Dry strip cover fluid was poured on to the cooling plate and the Drystrip tray with its anode at the upper end [red] and the cathode at the lower end [black] was placed on it. The cathode and anode ends were connected to the Multiphor II isoelectric focussing unit. About 15 mL of the Dry strip cover fluid was poured on the tray and the Dry strip tray aligner was placed on it with the groove side facing upwards. Care was taken to avoid formation of air bubbles while aligning the dry strip tray on the cover fluid since it could affect the thermal contact. The

rehydrated Dry strip gels from the reswelling tray were removed and placed parallel to each other along the lanes of the tray with their gel side facing upwards. The cathodic and anodic ends of the strips along with the gel surface were kept in contact by the moistened electrode strip. The electrodes were placed and pressed down on these electrode strips. The cover fluid was poured over the strips to immerse them completely and a programmed EPS 3501 XL power supply (Amersham Biosciences) was switched on. The Techne Circulator C-100 with a TECAM Heat exchanger was used to maintain a temperature of 20°C throughout the process of separation. The gels were left to run for atleast 8 h and the strips were stored at -80°C in plastic petridishes until use. The parameters used in running the IEF strips have been detailed in Tables 1 and 2.

7.7 Casting of SDS-PAGE gel

Two glass plates of size 20 X 20 cm and 20 X 22 cm were wiped clean with 50% ethanol. The glass plates were aligned parallel to each other with a spacer of 1 mm thickness placed in between them around the edges. Single screw clamps were used to clamp the plates together tightly and the whole set up was mounted on a rubber gasket placed on a casting stand.

A 12% (w/v) resolving gel was prepared as mentioned in the materials section and the solution was degassed for 30 min after which it was pipetted out into the space between the glass plates upto about 4 cm from the top of the smaller plate. The surface of the solution was overlayed with 18.2 MΩ/cm water and allowed to polymerise for 45 min atleast. After the gel was set, the overlayed water was poured out and a clean filter paper was used to remove any residual liquid.

A 4% (w/v) stacking gel acrylamide solution was prepared and layered on top of the resolving gel solution. A comb containing a long 17 cm well and a small size standard well measuring 1.5 mm in thickness was immediately inserted into the gap between the plates. Care was taken to avoid formation of air bubbles by inserting the combs at an angle. This was allowed to polymerise for 30 min at least. Once set, the combs were removed and the surface of the gel was rinsed with copious amounts of 18.2 MΩ/cm water to get rid of any gel debris or unpolymerised acrylamide.

The frozen IPG strips were equilibrated with 0.1 g of dithiothreitol (DTT) in 10 mL of the equilibration stock solution by gentle rocking on a shaker for 15 min. The equilibration step was repeated again with 0.45 g of iodoacetamide (IAA) instead of DTT. IAA was dissolved in 10 mL of the equilibration stock solution and the strips containing the gel were gently rocked on a shaker for 15 min.
The strips were placed on a filter paper moistened with 18.2 MΩ/cm water to drain all the excess liquid. Approximately 1 cm of plastic from the negative edge of the strip was cut off in order to obtain a suitable length. The strips were layered on the top of the gel and liquid agarose sealing solution was poured in between the glass plates allowing it to cool down. Care was taken to ensure that the strips were placed with their gel side facing the smaller glass plate.

The gels containing the strips were loaded onto the SDS PAGE electrophoresis unit by assembling them on a cooling core which was gently immersed at an angle into a gel tank containing about 2.0 L of running buffer. Protein size standards of 8 µl were loaded into the respective wells of each gel present on either side of the cooling core. Running buffer was also used to fill up the upper compartment containing the electrodes.

The second dimension was run in the electrophoresis unit where the lower and upper parts of the tank were filled with the running buffer. A current of 8 mA was applied to two gels and the process was normally complete in about 4 h which was indicated by the bromophenol blue dye that ran down the gel and started to come off. Throughout the separation process, the cooling core was maintained at a temperature of 10°C using the Techne Circulator C-100 with TECAM Heat Exchanger 1000.

On completion of the process, the clamps were removed and the gels were transferred to large storage containers and 250 mL of the fixing solution was added to it. The gels were fixed for about 1 h with gentle rocking.
The fixing solution was replaced by 250 mL of the Coomassie blue staining solution which was left rocking on a shaker overnight at room temperature.

Following this, the gels were destained using 18.2 MΩ/cm water for 4-6 h and the images were scanned and analysed using the PDQuest software. The raw image data obtained from scanning the gels using GS-800 Bio Rad densitometer was uploaded and used in analysis in the Redfin and SameSpots softwares.

7.8 Fluorescent staining

Attempts were also made to use fluorescent staining for visualising gels containing protein spots. Once the second dimension run was complete, the gels were left overnight in the fluorescent stain fixing solution with gentle rocking on a laboratory rocker. The gels were transferred to the staining solution the following day and allowed to stain from 3 to 8 h. The gels were stored in special storage boxes wrapped with aluminium foil to avoid contact with sunlight.

7.9 Visualisation of the stained gel

The Bio Rad Chemi doc XRS (Quantity One ™ software) was used to capture images of the gel using UV transillumination. Hard copies of the gel were printed out using the Mitsubishi Video Copy Processor attached to it.

7.10 Gram staining procedure

7.10.1 Heat fixed film preparation

A droplet of water was spread on to the surface of a clean glass slide with the help of a sterile plastic loop. A single colony of the bacteria was picked from a fresh culture plate with another sterile loop and mixed gently with the droplet of water and allowed to dry at room temperature. The slide was heat fixed by passing it through a Bunsen flame. Care was taken to avoid heating the slide too much since it could result in the charring of proteins. The slide was allowed to cool after heat fixation and subjected to Gram staining.

7.10.2 Gram staining

The heat fixed slide was flooded with methyl violet stain for 30 s. The stain was poured off after 30 s and the slide was flooded with Gram's iodine using a Pasteur pipette. The slide was left for 30 s with Gram's iodine on it and then gently rinsed off with tap water. Following this, the decolourising agent, acetone was used to flood the slide for just a few seconds (3-4 s) and washed off instantly in running tap water. The slide was finally flooded with Safranin stain for 30 s. The slide was rinsed off with tap water and dried with a blotting paper. Once the slide was completely dry, it was observed under the microscope using the oil immersion lens (X 100).

7.10.3 Expected Result

The presence of violet or deep purple stained cells indicate that the bacteria may be Gram positive and the presence of red, pink or orange cells indicate that they are Gram negative in nature.

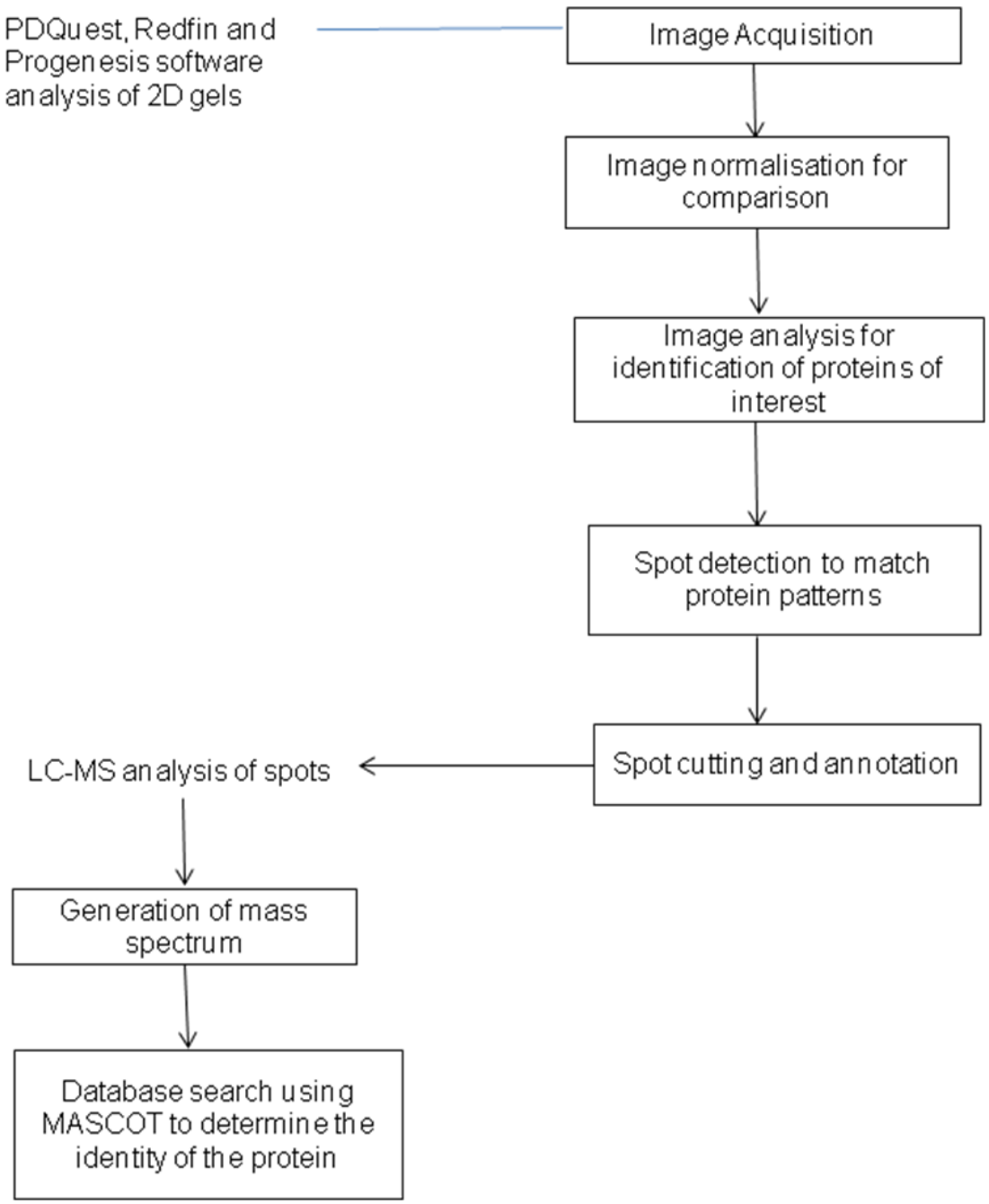


Figure 11: Methodology used for the analysis of spot data

7.11 PDQuest analysis

7.11.1 Acquiring images

The Bio-Rad GS-800 image densitometer was used for scanning and obtaining images of the 2D gels. Atleast 6 gel images were generated from every condition in order to ensure reproducibility of results. The cultures were grown in 450 mL volumes and separated out into 4 samples (Figs. 20.1-20.4, 21.1-21.4, 22.1-22.4, 23.1-23.4, 24.1-24.4, 25.1-25.4) of 100 mL volume each when they reached the appropriate optical density. A second lot of 450 mL cultures were grown independently to process the rest of the 2 or 3 samples (Figs. 20.5-20.7, 21.5-21.6, 22.5-22.8, 23.5-23.6, 24.5-24.6, 25.5-25.6)). Proteins were extracted from these biological replicates and 100 µL of the cell free extract from each sample was processed separately using the 2D Clean up kit. Hence 4 samples of the cell free extracts used in sequential optimisation experiments were from the same pool of material whereas the other 2 were from a separate culture. The PDQuest ™ software uses Bio-Rad imaging services to convert the signals from biological samples into digital data. The gel was initially 'Preview' scanned and specific sections of the gel which contained all the protein spots were framed and acquired. Fig. 9 shows the steps involved in the processing of data.

All the gels were scanned at the same time and care was taken to keep the frame dimension the same in all the cases. Once scanned, the images were saved and copies were printed out using the Mitsubishi Video Copy Processor. The gels can be cropped or rotated according to the crop and rotation tools on the Image menu. PDQuest was used as the default software for analysis based on the Student t-test.

7.11.2 Spot detection

The spot detection wizard was used to start up an experiment and step wise analysis of the gels was carried out. The spot detection wizard was used to detect spots on the gels. A template gel was selected and parameters were generated based on the smallest, faintest and the largest spot identified in it. The set parameters were used to scan every gel accordingly since gels show a high degree of variability with respect to streaking, intensity or clarity. The spot detection wizard was set to identify Gaussian, salt and pepper noise characteristics and filter them using a 7 X 7 pixel size filter dimension based on the individual noise pixels in the gel. Other features

including subtracting backgrounds and speckles detection were also enabled to make the detection process more efficient.

Once the gels were filtered and smoothed to clarify all the spots including fuzzy, streaked or overlapping spots, three dimensional spots called Gaussian spots were created. All types of analysis were then performed on the Gaussian image which contained the Gaussian spots that were precisely identified and quantified. Three types of images were created during spot detection namely the original 2D scan, the filtered image and the synthetic Gaussian image. The Gaussian spots were indicated on the gels using crosshairs that was located at spot centres and spot ellipses that encircled the spot shape.

7.11.3 Editing spots

Editing of spots involved several features including adding, removing, combining or changing spot boundaries. All the editing was performed on the Gaussian image. Since the same parameters were used to identify spots in all the gels, some spots seemed to have been missed out. In certain cases, air bubbles or stain marks were identified as spots. Hence all the gels were reviewed manually to minimise errors. The spot adding tool was used to add spots to the Gaussian image and the spot removal tool was used to remove incorrectly identified or positioned spots.

Other tools include the spot combining tool which is used to combine spots that have been identified to be different and the changing boundary tool which helps with the altering of spot boundary either by removing or moving them around.

7.11.4 Matching spots

Setting up an experiment automatically results in the generation of a matchset once the spots on the gels have been detected. Further grouping of the gels based on the various media conditions in which they were grown was also performed. The three main groups were the Media without mucin, Media with Type II mucin and media with Type III mucin. The gels could be matched manually or automatically. Automated matching was performed with a 50% restriction where the master image was matched to all the other images. The master image contained all the spots from all the gels mapped to it. However, a classic manual match was also performed after the automated match in order to minimise errors during matching. Landmarking was the first step performed which involved choosing reference spots to align and

position gels for matching. Once the landmarks are selected, an automatic match of all the spots is performed. The remaining spots can be reviewed individually and either matched or simply added to the master.

7.11.5 Data Analysis

Upon completion of the matching, an experiment summary report can be obtained from the analysis menu. The report shows the number of spots identified in each gel including its group details, match rates 1 and 2. Match rate 1 shows the percentage of matched spots relative to the total number of spots on the gel. Match rate 2 shows the percentage of matched spots on the gel relative to the total number of spots on the master gel.

Standard spot numbers (SSP) are assigned automatically to all the spots. These are generated by the software when the standard gel is divided into rows and columns with each section having the same number of spots. The spots in a section are numbered sequentially and the first two digits of the SSP number indicate the X and Y coordinates of that section. The spot review tool was used to study the histograms of all the spots in a matchset. The spot quantities, intensities, location and the number of gels showing expression of a protein can be identified using this tool. The highlighted spot can be compared between all the gels and match modifications can also be made. The histograms of interest can be selected and used to create analysis sets.

The scatter plot analysis shows the relatedness between two gels or two groups in a matchset. The quantity of each spot in the first gel is plotted along the X axis against the quantity of spots in the second gel. The slope of the scatter plot shows the degree of similarity. The closer the value is to 1.00, the stronger the relatedness between the gels.

From the reports menu, quantity and quality tables can be obtained where the spot quality report shows a range of values from 0 to 100 based on the Gaussian fit, X and Y streaking , overlap and linear range of the scanner.

7.11.6 Normalization

Normalization refers to the process where quantitative data from different gels are adjusted in order to compare different samples to one another. Variations in gels occur due to a number of factors which include handling and pipetting errors, loss of sample during loading, low cell density or inconsistent staining. Hence this process helps to compensate for non expression related variations in spot intensity.

The normalization formula:

Normalized spot quantity= Raw spot quantity X Scaling factor X Pipetting error compensation factor/ Normalization factor.

Where raw spot quantity refers to the intensity of each spot, scaling factor refers to a constant value defaulted as 106 parts per million (ppm) and the normalized quantity is multiplied to it.

The normalization factor was calculated for each gel based on 'total quantity in valid spots' method where the raw quantity of a spot in a gel was divided by the total quantity of all the spots in each gel included in the master.

All the spots which showed variations in expression in at least 6 gels were selected for analysis using LC-MS. Further, spots that showed expression in all three conditions were also selected for analysis.

7.11.7 Consensus

Group consensus was used to find potential spot detection errors and use the spot matching information uniformly across a set of gels in replicate group. The consensus tool helps to edit and match spots across specific gels, replicate groups or all the gels in a matchset. The ambiguous spots could be identified from the consensus analysis and matched to one or more spots on other gels. Once updated, the results are added to the spot review analysis report.

7.11.8 Gel analysis software

Apart from the PDQuest ^{TM} Advanced software, the Ludesi Redfin software and the Non linear Dynamics Progenesis SameSpots software were also used to compare and analyse the spot data. Spot data and match result tables have been attached to the Appendix F. Redfin software uses the Analysis of variance (ANOVA) and principal component statistical analysis (PCA) for summarising the spot data in *E.*

cancerogenus and *B. fragilis* whereas a cluster analysis is used to describe the same data in Progenesis Samespots software. Study of gels using three different software packages was performed to determine the reliability and reproducibility of analyses data from PDQuest, Redfin and SameSpots rather than depend just on the data obtained from PDQuest analysis. The differentially expressed spot identification criteria included a significant p value of 0.05 or less and a fold change of 1.0 or more.

7.11.8.1 Redfin analysis of 2DE gels

The six main steps used in the analysis of data using the Redfin software included the pre-processing of images, choosing of a warp reference image, warping of all the gel images, creation of a fusion image, detection of spots and generation of spot borders. Completion of all the above mentioned steps provided a series of results showing the experiment overview, statistics used, details of filtered spots including details on its p value (p values< 0.05 are considered to be significant), presence (set as 100%), volume and fold change with respect to all the three different growth conditions.

The pre-processing of images involved an initial quality control check on the gels based on a number of factors to determine its suitability for analysis. The image editor was used to further crop, flip or rotate images as necessary. The step involving the creation of a fusion image is automated and involves transferring of the detected spots and their borders on to every individual image for quantification. This step ensures 100% matching and helps to get rid of cracks or other damages to the image. The ‘’less-more’’ slider helps to optimise the number of spots detected on a gel and the ‘’loose-tight’’ slider helps to determine the spot boundaries.

Redfin analysis of spots also supports the categorisation of data into multiple groups to facilitate comparison. Once the reference image is selected and warped, global and local spot editing tools can be used to edit the spots on the gel. Once the spot borders are detected, an analysis report is generated containing the entire list of spots detected. The spots of interest are then filtered out by setting a p value filter of <0.05, a fold change of 1.0 and an expression volume of 100%. The filtered spots

could then be exported as a .pdf or .xls document. Spots of interest can be viewed using a 2D and 3D montage or statistically analysed.

7.11.8.2 SameSpots analysis of 2DE gels

The steps used in the analysis of data using the Progenesis SameSpots included the initial quality assessment followed by the alignment of images. The initial quality assessment step also helps to prevent positional errors with tools for cropping, flipping or rotating images. This software uses an automated image alignment system in order to minimise the repeated editing and re-matching of images.

The pre-filtering step of analysis involved the filtering out of damaged, noisy and spots from the edges of the images. This was followed by appropriate grouping of gels based on their growth conditions to study differential expression and an automated gel analysis involving spot detection, background subtraction, normalisation and matching was performed.

The analysis provides a review of results containing a list of spots from which the spots with relevant p values (<0.05) were selected. Fold change of 1.0 or more can also be used to select relevant spots since fold change is calculated by subtracting the highest and the lowest mean values and this is added on to the mean value of the third group. They provide a 2D and 3D montage view of spots to be able to study them better.

Apart from analysis of variance, this software also uses other statistical applications like PCA analysis, correlation analysis and power analysis for describing the data obtained from the experiment. PCA analysis helps to determine the variation in expression based on the levels of spot expression spread out across all the gels.

The correlation analysis helps to determine the degree of relatedness between two sets of data. Normalised spot expression volumes with a value closer to 1 indicate a higher correlation when compared to lower correlation spots which have a value closer to -1. The results of the correlation analysis are visually presented using dendrograms. Based on the variations in expression, difference between mean groups and sample size, the power value is calculated. The larger the difference between the group means, the more significant the power value of the spot. Usually a power value of 0.8 or more is categorised as a relevant spot. The data is presented

as average values based on the differential expression of spots in three different conditions of growth. The spots of interest could then be selected and used to generate a report for future use.

7.12 Preparation of protein spots for analysis using the Mass spectrometer

The spot outlines of interest were cut out using a scalpel and the gel bits were stored in silanised 0.5 mL tubes at -20°C until use.

The Coomassie blue stain was removed from the gel slice by adding 100 µL of 100 mM NH_4HCO_3 and 60 µL of acetonitrile and shaken for 30 min. The same washing process was repeated at least 3 times in order to get rid of the stain completely. Following this, the gel slices were dehydrated by adding 50 µL of ACN and incubated in room temperature for 5 min. The same process was repeated again. The gel slices containing the proteins were dried using the centrifugal evaporator for 15 min at room temperature. In the meanwhile a 20 µg/mL concentration of trypsin solution was prepared in 50 mM NH_4HCO_3 solution and 25 µL of the trypsin solution was added to each of the tubes. The tubes were incubated on ice for 30 min and 30 µL of 50 mM NH_4HCO_3 was added to cover the solution before incubating it at 37°C in a waterbath overnight.

The following day, 30 µL of a 50% (v/v) ACN and 5% (v/v) formic acid solution was added to stop the reaction and the tubes were shaken for 30 min. The solution containing the digested peptides were removed and stored in a clean microcentrifuge tube. A 30 µL solution containing 83% (v/v) ACN and 0.1% (v/v) formic acid was added to the tubes containing the gel slices and shaken for 30 min to remove all the remaining digested peptides which were then pooled into tubes containing the initial collection of digested peptides.

7.13 Liquid chromatography

Liquid chromatography experiments were performed using the Dionex Ultimate 3000 nano LC system which consists of a pump that regulates the flow of buffers, temperature compartment that is maintained at 60°C while running samples, the autosampler that is attached to the syringe facilitating injection of the samples from

the multi-well plate and the UV detector that shows the absorbance values of samples at 254 nm. Separation of samples was carried out using a 200 µm X 5 cm long monolithic column made up of polystyrene divinyl benzene polymers. The peptide elution gradient was set in such a way that the first 0-0.5 min was used to perform equilibration and injection at a 100% Buffer A concentration. The elution of peptides occurred between 0.5 and 16 min of the gradient where the concentration of acetonitrile was gradually increased by pumping Buffer B to a 100%. The column was washed and prepared for the loading of the next sample during the last 6 min of the gradient (Total 22 min gradient).

The nano LC system is coupled to a Hystar^{TM} Bruker Daltonics ESI Ion trap MS system which carries out the MS/MS fragmentation of peptides.

7.14 Mass spectrometric analysis

The microcentrifuge tubes containing the digested peptides were freeze dried completely and then resuspended in 10 µL of the start buffer A. The solution was then pipetted into the respective microtitre well for performing an LC-MS analysis.

The HCTUltra was used as the ion trap system with the ESI source. In order to create a stable spray with high ionisation yields, optimised ESI conditions were used. Identification experiments were performed at a flow rate of 3 µL/min with a nebulizer pressure of 15 psi and a drying gas flow rate of 5 L/min at a drying gas temperature of 300°C.

7.14.1 Tuning

Optimising the voltages that guide the flow of ions along the CapExit, Skimmer, Octopoles, Lenses and Tap Drive help to achieve maximum signal intensity called tuning. The Smart parameter setting was used for qualitative analysis and this was done using a standard compound. The target mass was set to 622.03 m/z. Tuning solution that was used in our experiments was 20% (w/v) acetonitrile.

7.14.2 Parameter settings for peptide mapping

The Standard enhanced scan mode was used to obtain MS spectra in the order of singly, doubly and triply charged precursor ions. The Ion Charge Control (ICC) target was set to 200.000. Scan range was 300-2000 m/z for MS and the average was set to 5 (Low concentration samples required higher average numbers whereas low average numbers were set to complex samples).

The Octopole 1 DC was set to 8.0 V. All ions above the intensity threshold were selected in the Auto MS (n) precursor selection analysis. Singly and doubly charged ions were detected whereas other mass ranges were excluded as being contaminants.

The compound spectra generated from the Auto MS (n) were exported to the Biotools software as *.mgf files and then database searched through Mascot.

Ion chromatograms from data processing included the UV chromatogram, BPC All MS and the TIC All MSn. The MS/MS ion search was used to connect to the Mascot search where the query file peptides were compared to the peptides present in the database to determine the identity of the protein.

The username, email and an appropriate search title were filled in the search box. The taxonomy used for carrying out the searches was ‘Eubacteria’ since the protein spots were cut out from gels expressing Gram negative bacterial culture proteins. The enzyme used was trypsin and the database normally used to conduct the searches was NCBI. The global modifications were set to a default of Carbamidomethyl (C) and the variable modifications were set to Oxidation (M). The missed cleavages were set to 2 and the charge state to doubly and triply charged ions. The mass tolerance was set to 1.7 Da, MS/MS tolerance to 0.5 Da and the output was set to 50 proteins. The results were saved by clicking on the ‘get hits’ option. (Refer to section G1 of the Appendix for more details) Random matching results could also provide false positive results and hence certain criteria were set up to segregate false positives from correct identifications. This included selection of results where the protein score was more than 40 and at least 2 peptides matched to the query sequence.

8. Materials used for the cloning, purification and structural studies of glycoside hydrolase enzymes from families 95 and 97 in *B. fragilis*

8.1 Liquid media

Unless stated otherwise, all media were prepared using distilled water, and sterilised by autoclaving. In cases where solutions could not be autoclaved, they were filter-sterilised using 0.2 µm Ministart® filter units (Sartorius) and stored in sterile plastic 30 mL containers.

Liquid media were stored at room temperature, and solid media at 4°C. All adjustments to the pH of solutions are stated under the list of components, and were achieved using HCl or NaOH.

8.2 Solid media

Columbia blood agar containing 5% (v/v) horse blood

Per 100 mL

Columbia agar powder	3.7 g
Horse blood	5 mL

To 95 mL of distilled water, the agar powder was added and mixed well before autoclaving. The medium was then allowed to cool to about 50°C and 5 mL of horse blood was added to it and mixed gently. This was aseptically poured into labelled Petri dishes and allowed to set for 30 min. The horse blood was aliquoted out in 5 mL quantities, stored in the freezer at -20°C and thawed before use.

8.3 Plasmid vectors and competent cells used for cloning

The plasmid vector used was pET-YSBLIC which was a modified form of the pET-28a vector developed in the University of York to allow ligation independent cloning (LIC) with over expression of proteins in BL21 strains. Refer to Appendix C for details.

The vector has the presence of a hexa-histidine tag at the *N*-terminal end that helps in easy purification of proteins using the immobilised metal affinity chromatography column.

The chemically competent cells used for the transformation were TOP10 *E. coli* cells and *E. coli* strain BL21 respectively.

8.4 Luria-Bertani (LB) broth

Per 1000 mL

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g

pH 7.0

8.5 LB agar

Per 100 mL

LB broth	100 mL
Agar (bacteriological agar N° 1)	2 g

The medium was autoclaved to make the agar soluble and then poured into Petri dishes when cooled to about 55°C.

8.6 Selective media

The antibiotic used in the selective medium was kanamycin where a stock concentration of 10 mg/mL was prepared in 18.2 MΩ/cm water and stored at -20°C in the freezer. A 1 in 100 dilution was used as the working concentration making it 50 µg/mL.

Isopropylthio-β-D-galactoside (IPTG) was used as an inducer in LB medium at a final concentration of 240 µg/mL and this was prepared from a stock concentration of 24 mg/mL dissolved in sterile 18.2 MΩ/cm water.

8.7 Cryogenic storage of bacterial stocks

The *B. fragilis* stock was stored at -80°C after mixing 0.5 mL of an overnight anaerobic basal broth culture with 0.5 mL of sterile 50% (v/v) glycerol.

Per 100 mL

100% (v/v) glycerol (Fisher)	50.0 mL

8.8 SOB broth

Per 100 mL

Tryptone	2.0 g
Yeast extract	0.5 g
5 M NaCl	0.2 mL
1 M KCl	0.25 mL
4 M $MgCl_2$-$6H_2O$	0.25 mL
4 M $MgSO_4$	0.25 mL

The solution was adjusted to a pH of 7.0 and autoclaved before use. $MgCl_2$-$6H_2O$ and $MgSO_4$ were prepared and autoclaved separately and then aseptically added to the solution.

8.9 SOC broth

Per 100 mL of the SOB broth

1 M D (+) glucose (filter sterilised)	2.0 mL

8.10 NZY Enrichment broth

Per litre

NZ Amine	10 g
Yeast extract	5 g
NaCl	5 g
$MgSO_4.7H_2O$	2 g

The solution was adjusted to a pH of 7.5 before autoclaving.

8.11 NZY supplement

Per 45 mL

1 M $MgCl_2$-$6H_2O$	12.5 mL
1 M $MgSO_4$	12.5 mL
20% (w/v) D (+) glucose	20 mL

The solution was filter sterilised and 0.45 mL of the NZY supplement was added to 10 mL of the NZY broth to make up the NZY^+ enrichment broth aseptically.

8.12 TE buffer

Per litre

0.5 M Tris-base pH 7.5	20 mL
0.5 M EDTA pH 8.0	2 mL

The buffer was adjusted to a pH of 7.5

8.13 Starter cultures for auto-induction media

Per 10 mL

Water	9.55 mL
1M $MgSO_4$	20 µL
1000 x metals	2 µL
40% (w/v) glucose	125 µL
25% (w/v) aspartate	100 µL
50 X M	200 µL
kanamycin	100 µL

8.13.1 ZYM-5052 auto-induction media

Per 500 mL

ZY	478.5 mL
1M $MgSO_4$	1 mL
1000 X metals	0.5 mL
50 X 5052	10 mL
50 X M	10 mL
kanamycin	5 mL

8.13.2 Stock solutions

8.13.2.1 50 X M

Per 100 mL

Na_2SO_4 (anhydrous)	3.6 g
NH_4Cl (anhydrous)	13.4 g
KH_2PO_4 (anhydrous)	17.0 g
Na_2HPO_4 (anhydrous)	17.7 g

The salts were sequentially dissolved in distilled water and autoclaved before use.

8.13.2.2 1000 X metals

Per 100 mL

Sterile 18.2 M Ω/cm water	36 mL
0.1 M $FeCl_3$-$6H_2O$	50 mL

(Dissolved in 100-fold dilution of HCl, approximately 0.12 M HCl and filter sterilised using 0.2 µm filters)

1 M $CaCl_2$ (anhydrous)	2 mL
1 M $ZnSO_4$-$7H_2O$	1 mL
0.2 M $CoCl_2$-$6H_2O$	1 mL
0.1 M $CuCl_2$-$2H_2O$	2 mL
0.2 M $NiCl_2$-$6H_2O$	1 mL
0.1 M Na_2MoO_4-$5H_2O$	2 mL
0.1 M Na_2SeO_3-$5H_2O$	2 mL
0.1 M H_3BO_3 (anhydrous)	2 mL

All the solutions were made separately and autoclaved unless stated otherwise.

20 mL of a 25% (w/v) aspartate solution was prepared and the pH was adjusted to 7.0 by neutralisation with NaOH and autoclaved separately.

8.13.2.3 50 X 5052

Per 100 mL (The compounds were sequentially dissolved in distilled water)

Glycerol	25 g
Glucose	2.5 g
α-lactose	10 g

8.13.2.4 ZY

Per litre (sequentially dissolved in distilled water)

Tryptone	10 g
Yeast extract	5 g

8.13.3 PASM-5052 SeMet-labelling auto-induction media

Per 500 mL

Water	450.3 mL
1 M $MgSO_4$	1 mL
1000 X metals	0.5 mL
50 X 5052	10 mL
20 X P	25 mL
17 amino acid mix	10 mL
25 mg/mL methionine	0.2 mL
25 mg/mL selenomethionine	2.5 mL
100 µM vitamin B12	0.5 mL
kanamycin	5.0 mL

8.13.4 Stock solutions

8.13.4.1 20 X P

Per 100 mL (sequentially dissolved in water)

$(NH_4)_2SO_4$ (anhydrous)	6.6 g
KH_2PO_4 (anhydrous)	13.6 g
Na_2HPO_4 (anhydrous)	14.4 g

8.13.4.2 17 amino acid mix (17 aa mix)

Per 100 mL (sequentially dissolved in water)

Alanine	1 g
Arginine-HCl	1 g
Asparagine-H_2O	1 g
Aspartate	1 g

Na glutamate	1 g
Glutamine	1 g
Glycine	1 g
Histidine-HCl-H_2O	1 g
Leucine	1 g
Isoleucine	1 g
Lysine-HCl	1 g
Phenylalanine	1 g
Proline	1 g
Serine	1 g
Threonine	1 g
Tryptophan	1 g
Valine	1 g

The amino acids were dissolved by continuous stirring and also heating with the Bunsen when necessary. The solution was made up to a pH of 7.0 by neutralisation with NaOH, filter sterilised using 0.2 µm filters and stored at 4°C until use.

8.13.4.3 Methionine and Selenomethionine solutions

20 mL of a 25 mg/mL solution of methionine solution was prepared and autoclaved before use.

5 mL of a 25 mg/mL L (+)-selenomethionine solution was made up just prior to use and filter sterilised using a 0.2 µm filter.

20 mL of a 100 µM solution of Vitamin B12 was prepared by filter serilisation using a 0.2 µm filter and stored at 4°C until use.

8.14 Protein extraction and purification solutions and buffers

8.14.1 Cell resuspension buffer in protein extraction

Per litre (pH 7.4)

HEPES	4.76 g
NaCl	29.22 g
imidazole	0.68 g

8.14.2 Nickel column purification resin

Fast flow chelating sepharose from GE Healthcare

8.14.3 Gel filtration purification resin

HiLoad 16/60 Superdex 200 prep grade (All the resins were stored in 20% (v/v) ethanol)

8.14.4 Buffers used for nickel affinity purification

8.14.4.1 Start buffer

Per litre (pH 7.4)

HEPES	4.76 g
NaCl	29.22 g
imidazole	0.68 g

8.14.4.2 Elution buffer

Per litre (pH 7.4)

HEPES	4.76 g
NaCl	29.22 g
imidazole	34 g

8.14.5 Buffers used for gel filtration purification

Per litre (pH 7.4)

HEPES	4.76 g
NaCl	11.9 g

8.14.6 Buffer exchange buffers

The proteins were concentrated using 30 kDa concentrators.

For CAH06598 and CAH09443

Per litre (pH 7.4)

HEPES (5 mM)	0.595 g

8.15 Crystallisation screens and reagents

The screen compositions for the various conditions used for the preparation of crystal trays have been included in the Appendix H.

8.16 Chemicals and solutions used for SDS-PAGE

8.16.1 Resolving gel (12% (w/v) acrylamide)

40% (w/v) solution (37.5:1 acrylamide: bisacrylamide)	3 mL
Buffer B	2.5 mL
18.2 MΩ/cm water	4.5 mL
10% (w/v) ammonium persulphate (APS)	50 µL
Tetramethylethylenediamine (TEMED)	10 µL

8.16.2 Buffer B

Per 100 mL

2 M Tris-base pH 8.8	75 mL
10% (w/v) sodium dodecyl sulphate (SDS)	4.0 mL

8.16.3 Stacking gel (4% (w/v) acrylamide)

40% (w/v) solution (37.5:1 acrylamide: bisacrylamide)	0.5 mL
Buffer C	1 mL
18.2 MΩ/cm	2.5 mL
10% (w/v) APS	30 µL
TEMED	10 µL

8.16.4 Buffer C

Per 100 mL

1 M Tris-base pH 6.8	50 mL
10% (w/v) SDS	4 mL

8.16.5 Running buffer for SDS-PAGE

Per litre

Tris-base	30.3 g
Glycine	144 g
SDS	10 g

The pH of the solution was adjusted to 8.8 before the addition of SDS and the running buffer was made up at a 10 X stock concentration which was diluted at 1:10 before use with distilled water.

8.16.6 SDS-PAGE Loading buffer

Per 10 mL

60 mM Tris-base pH 6.8	0.6 mL
50% (w/v) glycerol	5 mL
10% (w/v) SDS	2 mL
14.4 mM β-mercaptoethanol	0.5 mL
1% (w/v) bromophenol blue	1 mL

Stored at 1 ml aliquots at -20°C

8.16.7 SDS-PAGE solubilisation buffer

Per 10 mL

SDS-PAGE loading buffer	7.6 mL
urea	2.4 g

Stored at 4°C

8.16.8 Coomassie Blue staining solution

Per litre

Coomassie Blue R-250	1 g
glacial acetic acid	100 mL
methanol	450 mL

Coomassie blue gel staining solution can be filtered through a filter paper and re-used.

8.16.9 Destaining solution for gels

Per litre

glacial acetic acid	100 mL
methanol	100 mL

The Coomassie blue gel destaining solution can be filtered through a filter paper funnel containing activated charcoal and re-used once the stain is removed.

8.16.10 Bradford assay

Bradford's reagent solution	200 µL

Bovine serum albumin solution at various concentrations ranging from 0.1 mg/mL to 1 mg/mL.

Protein solution to be assayed	20 µL

8.17 Solutions used for enzyme assay

8.18 Stock solutions

20 mM solution of HEPES pH 7.4	25 mL
10 mg/mL solution of BSA	20 mL
10 mM 4-methylumbelliferyl α-D-glucoside	10 mL
10 mM 4-methylumbelliferyl α-L-fucoside	10 mL
50 mM 4-nitrophenyl α-D-glucoside	10 mL
50 mM 4-nitrophenyl α-L-fucoside	1 mL
50 mM 4-nitrophenyl β-D-maltoside	1 mL

10 mg/mL α-glucosidase (Recombinant protein purified from the study) 100 µL

10 mg/mL α-fucosidase (Recombinant protein purified from the study) 100 µL

8.19 Reagents for agarose gel electrophoresis

8.19.1 TAE Running buffer (50 X stock)

Per litre

Tris-base (ultra-pure)	242 g
17.51 M glacial acetic acid	57.1 g
0.5 M EDTA pH 8.0	100 mL

A 1: 50 dilution of this buffer is used for making agarose gels and for running gels in the electrophoresis tanks.

8.19.2 Bromophenol blue (6 X) sample loading buffer

Per 10 mL

Bromophenol blue	0.025 g
Glycerol	3.0 g

A 1 X dilution of the buffer was added to the sample before loading it onto a gel.

8.19.3 Size standards for agarose gel electrophoresis

The Bioline hyperladder I or NEB 1 kb ladder was used as a size standard after diluting to a concentration of 1 µg/12 µL. Bioline hyperladder I size standard was added at a concentration of 1 µg/5 µL before loading it on to the gel and stored at 4°C for up to 12 months. Refer to Appendix D for details.

8.20 Kits, reagents and enzymes for DNA extraction and purification

Quiagen miniprep kit was used for the extraction of genomic DNA.

8.21 Polymerase chain reaction

1 µL of the forward primer was mixed with 1 µL of the reverse primer and made up to 5 µL with sterile 18.2 MΩ/cm water.

The other reagents include KOD DNA polymerase enzyme, 1M $MgSO_4$, deoxy nucleotide tri phosphate (dNTP), KOD DNA polymerase buffer and DNA to be amplified.

9. Methodology used for the cloning, purification and structural studies of glycoside hydrolase enzymes from families 95 and 97 in *B. fragilis*

9.1 Bioinformatics analysis

The gene of interest was identified from GenPept in the NCBI sequence viewer v2.0 and the FASTA format of the sequence was obtained. The conserved and semi conserved residues of the sequence was determined using the ClustalW software (www.ebi.ac.uk/Tools/clustalw).

The expert protein analysis system (EXPASY) identification and characterisation tool was initially used for the determination of the molecular weight and pI of the proteins of interest. The presence of signal peptides and transmembrane helices in BF3763 and BF0855 was identified using the SignalP 3.0 server (http://www.cbs.dtu.dk/services/SignalP/), and the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/). The 'Translate tool' in Expasy was used to translate the gene sequence into amino acid sequence thereby determining the amino acids that are present in the protein (http://www.expasy.org/tools/dna.html)

The coding sequence (CDS) for the BF0855 protein from GH95 is

```
   1 ttaagggttc aactcgatcc attcaccggc ttgagtatct atgttcaaga tatccgtgcc
  61 ggataaagta tagttcttat tcggcttcaa tgtaccggga cgggccattt tcaacatgcc
 121 gcctttctcc gaatagatgc gtacccgatt gatttctcca ccttccatcc gggccgatac
 181 caggaatgca cccatggcgc gtaagttttc gaaggataca tctttccatt ctttcggtat
 241 tgccgggaag atgcgtatta cgcctgtgtg actttgcagc aacatttctt gtatgcctgc
 301 tgcaaaagcg aagttacctt cgagcgtgaa ggggcggtaa gtgaagcggg atttgccgct
 361 ttgggtctgg tcaccgttgg catggaaggt gttcttcagg cagaagcatt ccgcaaaggt
 421 tttcagggct tgtgcggcac cttcaccatc gaatgcacgg gctttcatat tggcgagcca
 481 actgtatgaa tatcctgtcc agtagtccgg ccctacttta tcgagtctct tcaaggtggc
 541 gcggatgatg tgttgtgact tttctccgtc gctccagtct atcagtccca gcggatggat
 601 agccatggca tgtgaaaaat gacggtgtga ttctttgtag ggatatcctt tggcgaaagt
 661 gaggcaacct tcttcgtcta tgtcgtaatc cggtagttgg gcttccaggg atgcccagtg
 721 tccggcttcg tcggccaggt tcagttcgtg tgccagttcg gaagtagcct taaagaggaa
 781 gtgcatcatt gccaggtcat aattggtcat gtcgctgaac cacgcttgca gggagttgtc
 841 gaatatttcg ggacttgaac tgaattccag tttacgtact ccttcggggg taacttctga
 901 aatttgctcc agatagattg ccacatcctt gatgaatgga taagcacgct ctttgaggaa
 961 agtgcggtct gccgaatatt tccattgcag atagaagtgt tgtgccagcc atgcagctac
1021 ggtttgcgac atggagtatt gtatccatcc tcccatgggt tctcctgtca gggtacaaac
1081 ccccggtata ttcatccctt cggtaccgaa atagcggcgg gtgtagcgtt tgtatgcatc
1141 ccgttggttc cacaacgtgt tcagatagcc cattccttcg gtcaggtgat tgcctgtata
1201 ggccggccag taacttagct gggtatttaa atcgtgatga tagtctccct tccagggcgg
1261 aagtttgccg ttatcggctg tccatacagc ctgcagggag atggggtatg agtgttcgcg
1321 ggtagtggaa ccgaatttat acatttcgtt ctgatactgt ttttgcagta cagaatcggg
1381 tagagtgatc gatgattgtg cccagtactt gtcccaatat tccaggtgtg cctgataatc
1441 atgctttaga ccgcgttgca gggcggcttc tgccttttcg gaagcttgct cgccggataa
1501 ggaagaggtg acgctccagg taccgtacaa ggtttcgcct tcttgcttcc aacagactgt
1561 tacgtcatac gaaaagtctc cgtatccctt ttgatggtag gttatttggt ttccttcacg
1621 aaccactttt ccttgttgat agcctaatcg gtgcagatct tgtcccgaaa ccggatcgag
1681 gctgccatcc ggcttcgttt tattgtatac aggagtgatg atggagggtt ctatcggggt
1741 tttcaggtta cggaatacaa accatccgat cggttcggta gcatgaacga aagtctgcat
1801 ttgtgttccg tctgcccagt ctgcttcaca taaggcattg ttcaggtaaa gtcttacctg
1861 agtcggagtg ccgatctgct ctaatggaaa ttcgatggca gctccgggaa tcttggacgg
1921 cgcgggattc atgtcataag gccagtccag ctttttctgt acgggcagat agttcttttg
1981 acgaatatgt tcttttaccc aggagaagcg gaagttatct cccgacagac tgtctaccgg
2041 acgtaaatcc cataaatcgg ttctgtccag tgacaggcgt agggtggagt cacgttgcca
2101 taccaaggca ccgacggtag catttcccaa gggcattgct tcgtcccacg accgggcaag
2161 tccggtgaat acaagatcac tctctgaagg agtgaccgga caatagtcat acgagttgca
2221 actgctccac agaccacaac agagtagtaa tagtagtttg attttcat
```

The CDS for the BF3763 protein from GH97 is

```
   1 atgaaaagaa aaatgatgtc cttattactc gcattggcgg taataagcgg aagtagcgtg
  61 tacgctaaag tgattgacgt aatgtctccc aacggagcca ttaaagtatc ggtagacatc
 121 aaggaccgga tttattattc ggtgtcctat gataatgacc agttattaaa agattgctat
 181 ctcaacctgc aactgcagaa tgagacgtta ggtacgaatc cccacttacg gagcaccaaa
 241 cgtggaacca ttgacgaaag tgtaaaacgt gaaatacctt tcaagaatgc gatcgtaaga
 301 aatcactgta ataccctgag aatgaatttc agcggaaatt atgccgttga attccgcgta
 361 ttcgacaatg gtatcgctta ccgttttgtg acagataaaa aaggagataa catcgtaatg
 421 ggggaagact tcgcaattaa ctttccaacc aattataaag ctcatctctc ccaaccggat
 481 ggctttaaaa cctcatacga atgcccatat actcatgtag ataccgaaaa gtatgctgct
 541 accgaccgca tgagttacct gcctgtattg atagaaacgg ataaagcata taaaatactg
 601 atatctgaag ccgacttatc cgattatccc tgtatgttcc ttaaaagtac cggtaagaac
 661 ggaatgcagt ctatttttcc caaagcacct ttagccttcg gagaagatgg tgaccgtagc
 721 ctcaagatta ccgaagaagc cgattacatt gccaagacgg acggcaaacg ttcattcccc
 781 tggcgcatga tggtgatttc gaaagaagac aaagaactga ttgaaaacga aatggtgtat
 841 aacctgtctg ctccttgtgt tcttgaagac tacagttgga tcaaaccggg acaagtgagt
 901 tgggaatggt ggcacgacgc acgcctctat ggggtagatt tccgttcggg tttcaatatg
 961 gattcctata aatactacat tgacttcgca tccaagttcg gtattcctta tatcatcatg
1021 gacgaaggat gggcgaaaaa cacacgtgat ccgtttaccc ccaatcccac catcaatctt
1081 accgaactga taaaatacgg aaaggaccgc aacgtaaaaa tcgtactttg gctgccatgg
1141 ctgactgtcg agaatcattt cgacctcttt aaaacatttg ccgattgggg catcgcagga
1201 gtgaagatcg acttcatgga ccgcagtgac cagtggatgg taaactatta tgaacgtgta
1261 gccaaagaag ccgccaagca taaactgttt gtagattttc atggtgcttt taaaccagcc
1321 ggacttgaac gcaaatatcc gaatgtgctt tcctatgaag gcgtattggg catggaacaa
1381 ggtggtaatt gcaaacctga aaacagcatt tatctgccct ttatgcgtaa tgccgtggga
1441 ccgatggatt tcactccggg ttcaatgatc tctgcacagc cggaagacaa ccgttccacc
1501 cgggccaatg ccatgggctc aggaacacgt gctttccaaa tggctctttt catcatcttc
1561 gaaagtggtc tgcaaatgtt agccgacaat ccggtttact attacagaga acttccctgt
1621 accgaattta tcacaagtgt tcccgtcacc tgggatgaaa ccaaggtcct ctatgccaaa
1681 gtaggtgaag cagtcgtcgt agccaaacga aaaggagaac agtggttcat cggaggtatc
1741 accggcaatc aaccacaaaa catcgagatc gacctcggat tcattccggc aggacaatca
1801 ttcacattaa cctcatttga agatggcatt aacgctgacc gtcaagcaat ggattacaag
1861 aaaaaggagt ctaccgtgaa caatcaaacc cgcatgacat tgaaaatggt acgcaacggg
1921 ggatgggccg gaacaattaa aatgaaatag
```

9.2 BLAST analysis

The similarity of the genes and proteins of interest with other sequences was determined using the BLASTP (for proteins) and BLASTN (for DNA) links in the National Centre for Biotechnology Information site (www.ncbi.nlm.nih.gov/BLAST).

9.3 Primer Design

The primers were designed using the Web Primer server (http://genome-www2.stanford.edu/cgi-bin/SGD/web-primer) with a default value of 18 for the minimum length of the primer.

PCR primers for the amplification of the glycoside hydrolase family GH97 BF3763 gene in *B. fragilis* NCTC 9343.

Forward primer sequence:

5'- CACCACCACCAC ATGAAAGTGATTGACGTAATGTCTCCCAACGG-3' (44 bp)

Reverse primer sequence:

5'- GAGGAGAAGGCGCGTTATTTCATTTTAATTGTTCCGGCCCATCCCCC- 3' (47 bp)

Refer to Appendix I for tables showing the conditions used for the amplification of the BF3763 gene from *B. fragilis*

PCR of the glycoside hydrolase gene GH95_BF0855 from *B. fragilis* NCTC 9343.

Primer sequences used for amplification

Forward primer

5'-CACCACCACCACATGTATGACTATTGTCCGGTCACTCCTTCAGAGAG-3'

Reverse primer

5'-GAGGAGAAGGCGCGTTAAGGGTTCAACTCGATCCATTCACCGGCTTGAGT-3'

The primer sequences were ordered from MWG Biotech, Germany and the Ligation independent cloning (LIC) specific ends were added to the respective ends of the primer sequences.

9.4 Genomic DNA extraction and purification in *B. fragilis* NCTC 9343

The DNeasy tissue kit was used for the extraction of DNA as per a standard protocol. A volume of 10 mL of LB medium was inoculated with a single colony of *B. fragilis* from a pure culture CBA plate. The medium was incubated overnight at 37°C anaerobically.

0.5 mL of the culture was centrifuged at 5000 x g for 10 min. The supernatant was discarded and the pellets were resuspended in 180 µL of Buffer ATL (Recipe patented). To this, 20 µL of proteinase K was added and vortexed well. The mixture was incubated overnight at 56°C in a water bath. The lysate was vortexed for 15 s and 200 µL of Buffer AL (Recipe patented) was added to it. The sample was mixed thoroughly and 200 µL of absolute ethanol (96-100%) was added to it. The sample was vortexed again vigorously and the mixture was pipette out into a DNeasy Mini spin column placed in a 2 mL collection tube. This was centrifuged at 14000 x g for 1 min. The flow through was discarded and the collection tube was replaced. 500 µL of Buffer AW1 (Recipe patented) was added to the Spin column and centrifuged at 14000 x g for 1 min. The flow through was discarded again and the 2 mL collection tube was replaced. 500 µL of Buffer AW2 (Recipe patented) was added to the spin column and centrifuged at 20000 x g for 3 min. The flow through and the collection tube were discarded. The DNeasy mini spin column was placed in a clean 2.0 mL microcentrifuge tube and 200 µL of Buffer AE (10 mM Tris-Cl, 0.5 mM EDTA pH 9.0) was pipetted out onto the column. The sample was incubated in room temperature for 1 min and the DNA was eluted out by centrifuging at 14000 x g for 1 min. The eluted DNA of interest can be stored at -20°C until use.

Agarose gel electrophoresis was performed to confirm the presence of the purified *B. fragilis* genomic DNA.

9.5 Agarose gel electrophoresis

A 1% (w/v) agarose gel was prepared by adding 1 g of agarose in 100 mL of 1 X TAE buffer and bringing it to boil to dissolve the solid bits. This was allowed to cool to about 60°C and then poured into a mini gel casting tray. Care must be taken to avoid air bubbles. A 12 toothed comb was then inserted into the solution. The gel was then left to set on the bench for 20 to 30 min.

Following this, the gel tray was placed horizontally in the electrophoresis tank and 1 x TAE buffer was poured to submerge the gel completely.
The comb was removed without damaging the gel and the samples were loaded into the wells. The samples were prepared by adding 5 µL of amplified DNA to 1 µL of the loading dye (bromophenol blue 6 X) along with 5 µL of the Bioline hyperladder standard (Refer to Appendix D). The electrophoresis was carried out at 100 mA, 200 V for 35 min.

The gels were then visualised using the Bio-Rad Gel doc system (using Quantity One ™ software) and hard copies of the gel were printed out using the Mitsubishi Video Copy Processor attached to it.

9.6 PCR amplification of the genes of interest

KOD Hot start DNA polymerase was used to amplify the genes of interest in the PCR protocol using the Eppendorf MasterCycler™ machine.

9.7 Reaction components

PCR components added to each tube	GH95 (BF0855)	GH97 (BF3763)
Sterile 18.2 MΩ/cm water	34 µL	33.6 µL
12.5 µM Forward primer	1 µL	1.2 µL
12.5 µM Reverse primer	1 µL	1.2 µL
KOD reaction buffer 10 X	5 µL	5 µL
(Refer to Appendix for details of the KOD reaction buffer recipe)		
B. fragilis genomic DNA (50 ng/ml)	1 µL	1 µL
$MgSO_4$ (25 mM)	2 µL	2 µL
dNTPs (5 mM)	5 µL	5µL
KOD DNA polymerase (1U/µL)	1 µL	1 µL

The PCR conditions used to amplify the BF3763 and BF0855 genes in *B. fragilis* has been detailed in tables in the Appendix section I.

The amplified PCR products of BF3763 and BF0855 were analysed using agarose gel electrophoresis and purified using the MinElute PCR Purification Kit, QIAGEN.

To 90 µL of the PCR product, 450 µL of Buffer PB (Recipe patented) was added and mixed so that they were in the ratio of 1:5 respectively. The mixture was applied onto a MinElute column that was placed in a 2 mL collection tube and centrifuged at 14000 x g for 1 min. The flow through was discarded and the collection tube replaced. 750 µL of Buffer PE (Recipe patented) was used to wash the column and then centrifuged at 14000 x g for 1 min. The washing step was repeated again. The flow through was discarded and the column was placed in a clean 1.5 mL microcentrifuge tube. 10 µL of buffer BE (Recipe patented) was pipetted onto the centre of the membrane in the column, left for 1 min at room temperature and centrifuged at 14000 x g for 1 min. This step eluted the purified PCR product which was stored at -20°C until use.

9.8 Insertion of LIC T4 polymerase (T4 pol) reaction

The purified PCR product (10 µL) was mixed with 2 µL of 10 X T4 pol buffer (Refer to Appendix A2 for recipe details), 2 µL of 25 mM dATP, 1 µL of 100 mM dithiothreitiol (DTT), 0.4 µL of 2.5U/µL of LIC T4 DNA pol (Novagen) and made up to 20 µL by adding 5.6 µL of sterile 18.2 MΩ/cm water. The mixture was incubated at 22°C for 30 min. To stop the reaction of T4 DNA polymerase, the mixture was incubated at 75°C for 20 min and centrifuged at 14000 x g for 10 s.

For effective annealing of the genes, BF3763 and BF0855 to the LIC vector pET-YSBLIC, 2 µL of the reaction mixture resulting from the above was added to 1 µL of the prepared pET-YSBLIC vector (50 ng/µL) and incubated at 20-22°C for 10 min. This was then made up to 4 µL by adding 1 µL of 100 mM EDTA and incubated at room temperature for 10 min.

9.9 Preparation of electrocompetent TOP10 *E.coli* cells

The glycerol stock of the cells was used to inoculate a plate that was incubated overnight at 37°C. A single colony from this was used to inoculate 50 mL of low salt LB which was incubated at 37°C in a shaker at 200 rpm until it reached an OD of 0.5 to 0.7 at 600 nm. The culture was incubated on ice for 30 min and centrifuged at 4000 x g for 15 min at 4°C. A volume of 25 mL of cold filter sterilised 18.2 MΩ/cm water was used to resuspend the pellets which were then centrifuged again and the supernatant was discarded. The pellets were resuspended in 12 mL of cold 18.2

MΩ/cm water. The resuspension and centrifugation steps were repeated again reducing the volume of water to 5 mL and then to 2 mL. The steps were repeated once final time reducing the volume of water used to resuspend the pellets to 0.2 mL. The cells were aliquoted out into 40 µL quantities and stored at -80°C.

9.10 Transformation

A volume of 2 µL of the pET-YSBLIC vector containing the gene of interest was added to 40 µL of TOP10 *E. coli* cells, gently mixed with a pipette tip and incubated on ice for 1 min. This was transferred to a 2 mm electroporation cuvette which was kept ice cold and sterile. The cuvette was tapped gently in order to make sure that the solution covered the bottom of it. The cuvette was placed in a Bio-Rad ShockPod after drying the outside thoroughly and the lid was closed. The cells were electroporated using the BioRad Gene Pulser Xcell at 2.5 V for 20 ms. 1 mL of the SOC medium was added to the electroporated cells immediately and transferred into a clean sterile micro centrifuge tube using a sterile glass pipette. This was incubated at 37 °C for 60 min at 200 rpm and then plated out onto an LB agar plate containing kanamycin at a concentration of 100 µg/mL. The transformed cells were allowed to grow by incubating the plates at 37°C overnight.

9.11 Screening of colonies for successful inserts

The number of colonies present in the LB agar plates containing kanamycin was counted the following day. The colonies of interest were transferred onto a fresh 24 grid LB agar plate supplemented with an appropriate concentration of kanamycin. The first colony was streaked across the first grid on the agar plate. The cells remaining on the pipette were gently and thoroughly mixed into 50 µL of sterile 18.2 MΩ/cm water contained in a 0.2 mL PCR tube. The PCR tubes containing the cell-water suspensions were incubated in the PCR machine at 95°C for 5 min to disrupt the cells and centrifuged at 14000 x g for 2 min. The pellet was discarded and the supernatant was transferred into a clean sterile 0.2 mL PCR tube. This was stored at -20°C until use. The same process was repeated for each and every colony of interest on the agar plate.

9.12 PCR screening to identify colonies containing gene inserts

A volume of 10 µL of the supernatant from above was mixed with 1µL of the forward primer (0.4 µM), 1 µL of the reverse primer (0.4 µM), 5 µL of dNTPs (2 mM), 1 µL of (5 U/µL) Taq polymerase, 5 µL of Taq buffer (10 X) (Refer to Appendix A2 for recipe details) and made upto 50 µL with sterile 18.2 MΩ/cm water.

PCR conditions for the amplification of the BF3763 gene from transformed *E. coli* to screen for colonies containing successful inserts has been included in the Appendix section I.

The presence of successful inserts was confirmed by agarose gel electrophoresis. Refer to Methods section 9.5 for details on how to perform agarose gel electrophoresis.

Colonies containing successful inserts were used to inoculate LB broth media supplemented with an appropriate quantity of kanamycin and incubated overnight at 37°C in a 200 rpm shaking incubator.

9.13 Purification of the plasmid DNA from the colonies containing inserts

The Bio/Spin-PTM Plasmid Miniprep Purification protocol was used to purify the plasmid DNA from the *E. coli* TOP10 cell colonies containing the gene of interest. A volume of 1.5 mL of the cell culture from overnight incubation was centrifuged at 14000 x g for 1 min and the supernatant was discarded. The same step was repeated twice and the supernatant was discarded with a sterile pipette tip. A volume of 200 µL of Buffer I (Recipe patented) was added and vortexed to resuspend the cells. To this, Buffer II (Recipe patented) of a volume of 200 µL was added and mixed by inverting 5 to 6 times. To this, 200 µL of Buffer III (Recipe patented) was added and mixed by inverting 5 to 6 times and centrifuged at 14000 x g for 5 min. A spin column was placed in a collection tube. The mixture was pipetted into the column and centrifuged at 14000 x g for 30 s. The flow through was discarded and 700 µL of the washing solution was added to the column and centrifuged at 14000 x g for 1min. The flow through was discarded and the column was centrifuged at 14000 x g for 3 min to remove any residual ethanol from the washing step. The flow through was discarded and the column was placed in a clean sterile 1.5 mL microcentrifuge tube. A volume of 50 µL of TE solution was added to the column and

centrifuged at 14000 x g for 1 min to elute the plasmid DNA which was stored at -20°C until use.

9.14 Method used to prepare chemically competent cells

The glycerol stock of the cells was used to inoculate a plate that was incubated overnight at 37°C. A single colony from this was used to inoculate 50 mL of low salt LB which was incubated at 37°C in a shaker at 200 rpm until it reached an OD of 0.35 to 0.4 at 600 nm. The cells were incubated on ice for 30 min and then centrifuged at 4000 x g for 10 min at 4°C. After discarding the supernatant, the pellets were resuspended in 15 mL of cold 80 mM $MgCl_2$, 20 mM $CaCl_2$ solution and then centrifuged at 4000 x g for 10 min at 4°C. The pellets were then resuspended in 1 mL of 100 mM $CaCl_2$. The cells were incubated on ice for 1 h. The cells were added to 50 % (v/v) glycerol making it up to a final concentration of 15 %(v/v) before freezing it in liquid nitrogen and storing at -80°C until use.

9.15 Restriction digests of the recombinant BF3763 and BF0855 genes from pET-YSBLIC

The restriction enzymes used to carry out the digestion was *Eco* RI and *Xba* I according to the analysis of the sequence in Webcutter 2.0 server (http://www.firstmarket.com/cutter/cut2.html). The restriction enzyme digestion was carried out to remove the recombinant genes from the vector. The restriction enzymes were kept on ice and centrifuged for 3 s before use. To 5 µL of the plasmid DNA, 0.5 µL of *Xba* I, 0.5 µL of *Eco* RI, 1 µL of 10 X NE Buffer 2, (Refer to Appendix A2 for recipe details) 1 µL of BSA (50 µg/ml (w/v)) and 2 µL of sterile 18.2 MΩ/ cm water was added to make it up to 10 µL. The mixture was incubated at 37°C for 2 h and the presence of the recombinant genes was analysed using agarose gel electrophoresis (Refer to section 9.5 in the Methods section).

9.16 Transformation of the recombinant genes into BL21 cells

Chemically competent BL21 cells were prepared and 50 µl of the cells were treated with 2.5 µL of the purified plasmids by gentle mixing on ice for 5 min with a sterile tip. The cells were subjected to heat shock at 42°C in a water bath for 90 s and then immediately transferred back onto ice. A volume of 200 µL of NZY^+ medium was

added to the mixture aseptically. The cell wall recovery was induced by incubating the mixture in a water bath at 37°C for 45 min. The transformed cells were plated out onto LB agar plates containing an appropriate amount of kanamycin and incubated overnight at 37°C. The number of colonies showing successful inserts was 2 for GH95 and 3 for GH97.

The genes for GH95 and 97 have been cloned by previous students. All the details are available in the dissertations ('Isolation and transformation of GH95_BF0855 gene from *Bacteroides fragilis* NCTC 9343 in competent cells to produce soluble protein' by Cheun Hong Yeap, 2006 and 'Cloning and expression of glycoside hydrolase family 97 BF3763 (putative exported protein) gene from the *Bacteroides fragilis* NCTC 9343 ATCC 25285 in *Escherichia coli*' by Lee Ling, Bong in 2006). Only experiments pertaining to the purification and characterisation of these enzymes have been performed by me.

9.17 Solubilisation buffer experiment

9.17.1 Growth of bacteria in starter cultures

E. coli cells were grown in 30 mL sterile plastic containers containing 10 mL of LB media or 10 mL of autoinduction media starter cultures supplemented with an appropriate amount of kanamycin. The cultures were inoculated from an agar plate or glycerol stock using a sterile inoculation loop and grown overnight in a shaking incubator at 37°C and 200 rpm.

9.17.2 Growth of bacteria for expression and purification using IPTG induced media

An initial study was done to compare the expression of protein in LB and auto-induction media. In both cases, 0.5 mL of the starter culture of BL21 *E. coli* cells was used to inoculate 50 mL of the starter culture. Following the inoculation of the media, the cultures were grown at 37°C in a shaking incubator at 200 rpm.

When the culture reached an OD of 0.6, the flask containing LB medium was induced for the expression of the proteins of interest by the addition of 100 mM IPTG

(Isopropyl β-D-1-thiogalactopyranoside) to make the final concentration 1 mM and then the temperature was reduced to 30°C and the shaking reduced to 100 rpm. However, for auto-induction media cultures, the temperature was reduced to 30°C and the shaking to 100 rpm when it reached an OD of 0.6 without the addition of IPTG. The cultures were then incubated overnight and processed the following morning.

9.17.3 Small scale protein extraction for the solubilisation buffer experiment

The overnight culture was taken the following day and 3.0 mL of it was transferred into two clean microcentrifuge tubes of 1.5 mL each. The tubes were centrifuges at 14000 x g for 1 min. Following centrifugation, the supernatant was discarded and the pellets were dissolved in 300 µL of the start buffer. The mixture was ultra-sonicated for 5 s at amplitude of 15 and maintained on ice while doing so. The cells were centrifuged at 14000 x g for 10 min and 200 µL of the supernatant also known as the cell free extract was transferred into a clean microcentrifuge tube. The remaining supernatant was discarded and the pellets were resuspended in 300 µL of the solubilisation buffer. An SDS-PAGE gel was run with 20 µL of the sample and 5 µL of the loading buffer except the solubilisation mixture which was placed in a boiling water bath for 1 min to dissolve the pellets and 20 µL of it was loaded on to the SDS-PAGE gel. Molecular weight size standards were prepared with 8 µL of the loading buffer and run along with the samples in order to determine the size of the expressed protein.

9.17.4 SDS-PAGE electrophoresis

The BioRad Mini Protean III kit was used to perform 2D SDS-PAGE. Two clean glass plates of dimensions 10.1 X 7.2 cm and 10.1 X 8.2 cm were aligned parallel to each other and clamped. The larger glass plate had a spacer ridge of 0.75 mm attached to it. The two glass plates along with their casting clamp stand were mounted on the surface of a rubber gasket by applying vertical downward pressure. The resolving gel solution was made up in a plastic container and pipetted into the space between the two plates. A 2 cm mark was made from the top of the smallest plate and the solution was filled up to it. For efficient polymerisation, the surface of the gel was layered with 18.2 MΩ/cm water and the gel was allowed to polymerise

for 20 to 30 min. The water was removed from the surface of the set gel and blotting paper was used to remove any residual unset gel or water. Care was taken to avoid touching the surface of the gel with the blotting paper. The stacking gel solution was then made up and pipetted on to the surface of the resolving gel. This was filled up to the top of the small glass plate and a 10 toothed comb was inserted into the gap between the plates immediately (Size of the comb was 1.1 X 0.75 cm). The gel was then allowed to polymerise for 20 min and the combs were slowly removed to reveal the wells. Care was taken to avoid the formation of air bubbles while doing so and the surface of the gel was rinsed with 18.2 MΩ/cm water. The gels were placed in the electrophoresis module in such a way that the smaller glass plates faced towards the centre. The module was then lowered into the electrophoresis tank and SDS-PAGE (1 X) running buffer was poured into it to fill both the central and the outer compartments. Care was taken to ensure that the electrodes were completely immersed in the buffer and no air bubbles formed. The buffer in the central compartment was allowed to flow into the wells and formed a layer on the top of the gel.

The prepared samples were placed in a boiling water bath for 3 min and then centrifuged at 14000 x g for 1 min before loading them onto the gel with the help of a Hamilton syringe. The electrophoresis was performed at 120 mA and 200 V for about 50 min. The completion of the process was indicated by the migration of the bromophenol blue dye to the bottom of the gel.

Upon completion of electrophoresis, the gels were removed from the set up and placed in a plastic container containing Coomassie blue staining solution and rocked gently on a shaker for 10 min at room temperature. Following this, the gels were immersed in the destaining solution for atleast 6 to 8 h. Once the bands were clearly visible, the gels were rinsed with 18.2 MΩ/cm water and then visualised using the BioRad gel documentation system. Photographs of the gel were obtained using the Mitsubishi Video Copy Processor containing the Mitsubishi thermal paper (Refer to Appendix B for details).

9.18 Large scale growth of bacteria

A large scale growth of bacteria was carried out using the auto-induction medium where 500 mL of the medium was inoculated with 5 mL of the starter culture.

The growth procedure was followed exactly the same way as mentioned above.

9.18.1 Protein extraction

The induced cells were centrifuged at 4000 x g for 15 min and were resuspended in 5 mL of re-suspension buffer. The cells were ultra-sonicated on ice at amplitude of 14 for 2 min in 10 s bursts with 10 s intervals. The lysate was then transferred into centrifuge tubes and then spun at 24000 x g for 30 min at 4°C. The liquid supernatant or the cell free extract was transferred into a sterile plastic universal and stored in ice until purification.

9.18.2 Protein Purification

The recombinant proteins were purified using immobilised metal affinity chromatography (IMAC). Proteins have the presence of hexa-histidine tags at the N terminal that could be used to separate it out from other *E. coli* proteins using the Sepharose^{TM} chelating fast flow resin which acts as the metal affinity matrix. The column was first stripped and charged using 1M $NiSO_4$ followed by equilibration with 200 mL of start buffer. After equilibration, the cell free extract (10-20 mL) was loaded onto the column and then eluted out with 100 mL of the elution buffer at a flow rate of 4 mL/min. The presence of the protein was monitored using a UV spectrophotometer and the eluted fractions were collected in 4 mL quantities in plastic tubes. A volume of 20 µL of each sample was run on an SDS-PAGE gel to confirm and detect the fractions containing the pure protein. The pure protein fractions were then concentrated using a Vivaspin protein concentrator and then resuspended in the gel filtration buffer.

Further purification of the proteins was carried out using gel filtration chromatography. The column was first equilibrated using 120 mL of the gel filtration buffer at a rate of 1 mL/min. A volume of 0.5 mL of the protein sample was loaded onto the column by injection and then eluted out at a rate of 1 mL/min. The presence of the pure protein was detected using the UV spectrophotometer and the eluted fractions were collected in 5 mL quantities in plastic tubes. A 20 µL sample from every fraction was run on an SDS PAGE gel to confirm the presence of the pure protein. The fractions of interest were pooled and concentrated.

9.18.3 Concentration of proteins

The pure proteins were concentrated using the 30 kDa Vivaspin 6 mL cut off concentrator units (Viva Science). The pooled fractions were concentrated by centrifuging at 4000 x g and were buffer exchanged with 5 mM HEPES (pH 7.4) within the concentrator units.

9.19 Concentration of proteins determined by Bradford's assay

Standard dilutions of BSA were prepared within a range of 1-10 µg/mL and the total volume was made up to 500 µL. The absorbance of the solutions was recorded at 595 nm using a visible spectrophotometer and a standard curve was drawn based on the values obtained. A blank was set up containing 18.2 MΩ/cm instead of BSA and added to the Bradford reagent. The protein solution was then diluted appropriately and 500 µL volumes were transferred into cuvettes. The absorbance values were measured and plotted against the standard to determine the protein concentration.

From the standard curve, the value of c was found to be 0.6 which was the point on the y axis where the line intercepted. Substituting the rest of the values in the equation below, the value of m was determined to be 0.52.

Unknown concentrations of protein samples and their dilutions 1:10, 1:50 and 1:100 were prepared and mixed with the Bradford reagent. The absorbance values at 545 nm was measured and applied to the equation $y = mx + c$ where y = Absorbance at 545 nm, x = the concentration of the unknown sample, m = the gradient of the straight line and c = the point of intercept of the line with the y axis.

Once the concentration of the unknown protein was determined, it was multiplied by the dilution factor in order to obtain the concentration in mg/ mL.

Protein conc. mg/ mL= Conc. obtained from the curve X dilution of the sample/1000.

20 µL of known concentrations of BSA were prepared and mixed with 200 µL of the Bradford reagent and the absorbance was measured at 545 nm. The concentration range of BSA was 0.1 to 1.0 mg/mL to generate a standard curve and the standard curve has been included in the Appendix section J.

The concentration of the GH95 protein was determined to be 43 mg/mL from the equation with an absorbance of 0.823 and a dilution factor of 100. The concentration of the GH97 protein was determined to be 29.2 mg/mL from the equation with an absorbance of 0.752 and a dilution factor of 100.

The proteins were diluted to obtain a final concentration of 15 mg/mL with 5 mM HEPES buffer pH 7.4 before crystallisation.

9.20 Crystallisation

The protein concentration was determined using Bradford assay and then crystallisation trays were set up.

9.20.1 Crystallogenesis

1 ml of liquid Aquasil was mixed with 999 mL of 18.2 MΩ/cm water in a 2000 mL beaker and the cover slips were added to the solution.

The liquid solution was stirred every 15 min for about 2 h. The cover slips were taken out and allowed to dry on a tissue paper. The cover slips were wiped down with a clean cloth before use. The crystal tray wells were filled with the appropriate buffer solutions from various screen aliquots and the cover slips containing drops of protein and buffer in three different concentrations of 1:1, 1:2, 2:1 respectively were inverted onto the surface of the wells which were greased beforehand. Once the tray of 24 wells was set up, they were incubated at 22.4°C in the crystallisation incubator and monitored for the presence of crystals every week. The various crystal screens and their compositions have been included in Appendix H. Crystals were harvested in rayon fiber loops, bathed in cryo-protectant solution prior to flash freezing in liquid nitrogen.

9.20.2 Data collection

Diffraction data were collected on the European Synchrotron Radiation Facility (ESRF) from single crystals at 100 K with an oscillation range of 0.5 °. Data were collected at a wavelength of 0.933 Å on beamline ID14-2 using an ADSC Q4R CCD detector. All diffraction data were processed with MOSFLM and reduced and scaled with SCALA (Leslie, 1992). All other computing was undertaken using the CCP4

suite (Anon, 1994) unless otherwise specified. The structure of *Bf*GH97 was solved by molecular replacement using the program PHASER using the *Bacteroides thetaiotaomicron Bt*GH97b (UniProtKB/TrEMBL entry Q8A6L0) structure (Okuyama *et al*., 2009) (PDB code 3A24) as a search model. 5 % of the total reflections were set aside for cross validation ("R_{free}") purposes and for the weighting of geometrical and temperature factor restraints. The structure was manually built with cycles of REFMAC (Murshudov *et al*., 1997). Solvent molecules were added using COOT (Emsley & Cowtan, 2004) and checked manually. The structure was validated using MOLPROBITY (Davis *et al*., 2007).

9.21 Enzyme assay

9.21.1 Preparation of substrates

A volume of 1 mL of 6 mM substrates were prepared by dissolving the substrate in 18.2 MΩ/ cm. Since the methylumbelliferone substrates were found to be difficult to dissolve, they were ultra-sonicated (15 amplitude) at 1 min intervals until the substrates dissolved completely. The appropriate volumes required for the assay were aliquoted out into clean microcentrifuge tubes. Care was taken to perform the assays with freshly prepared substrates.

9.21.2 Enzyme assay protocol

Two different assay methods were used to characterise and determine the specific activity of the enzymes expressed from GH95 and 97 namely fluorogenic and chromophoric assays. The putative enzymes expressed from GH95 and 97 were α-fucosidase and α- glucosidase respectively.

9.21.3 Chromophoric assay

p-Nitrophenyl α-D-Glucoside →α-D-Glucose + p-Nitrophenol

CONDITIONS: T = 37°C, pH = 6.8, A 400nm, Light path = 1 cm

The substrates used were p-nitrophenyl α-D-glucoside, p-nitrophenyl α-D- fucoside, p-nitrophenyl ß-D-maltoside and p-nitrophenyl α-D-galactoside. The activities of the enzymes were measured by their action against the substrates resulting in the release of 4- nitrophenol indicated by a change of colour to yellow. The assay

mixture contained 20 mM of the substrate, 5 mM HEPES pH 7.4, 10 mg/mL BSA and 15 µL of the enzyme at a final concentration of 10 mg/mL. The volume of the assay mixture was made up to a total volume of 500 µL. The mixture was transferred to a clean plastic cuvette, mixed well with a sterile pipette tip and incubated at 37°C. The change in absorbance values at a wavelength of 420 nm was noted at 1 min intervals.

9.21.4 Fluorimetric assay

The substrates used to carry out the fluorimetric assays include 4-methylumbelliferyl α-D glucoside, 4-methylumbelliferyl α-D-fucoside and 4-methylumbelliferyl ß-D-maltoside. These are synthetic substrates that react with enzymes to release a product called 4-methylumbelliferone (4-MU). 4-MU is a fluorescent dye whose fluorescence was measured using the Bio-Tek FL600 fluorescence microplate reader. The excitation wavelength used was 360 nm and the emission wavelength was 460 nm. Standard assays were performed at 37°C in a total volume of 500 µL of 5 mM Hepes buffer, pH 7.4, containing 10 mg/mL BSA and 0.15 µg of the enzymes, α-D-glucosidase and α-D-fucosidase. The reaction mixture without the enzyme was transferred into a clean glass cuvette and placed into the cuvette holder in the fluorescence densitometer. Once the excitation and emission wavelengths were set, the enzyme was quickly added and mixed well with a sterile tip and the photometric values were measured at 10 min intervals for 60 min. The cell containing the reaction mixture was maintained at 37°C by a temperature controlled circulator. The photometric recordings were traced out onto chart paper using a chart recorder with a recorder pen attached to it. Positive activity was detected in the case of both enzymes.

Two different types of substrates were used in the study to determine the activity of the recombinant enzyme, α-glucosidase purified from our experiment. Para nitrophenyl- α-D- glucopyranoside was used as a substrate where α-D-glucose and p-nitrophenol were expected to be produced as end products as a result of the hydrolysis of the α 1→6/ α 1→4 linkages in the substrate. The release of para nitrophenol into the reaction mixture would result in the development of a yellow colour. The intensity of the yellow colour indicated the amount of product released and was measured spectrophotometrically at 420 nm in the colorimetric assay

(*Anon*, 1996). The other substrate that was used in the study was 4-methylumbelliferly α-D-glucoside. The α-glucosidase carried out an enzymatic hydrolysis of the fluorescent sugar resulting in the release of 4-methylumbelliferone as a byproduct of the reaction. The presence of the product in the reaction mixture was analysed using a fluorescent spectrophotomer where the excitation and emission wavelengths were used to measure the amount of product released in the fluorimetric assay (Sheldon *et al.*, 2006).

The BF0855 gene from *B. fragilis* NCTC 9343 was cloned and expressed in *E. coli* and the putative activity of the enzyme was predicted to be α-L-fucosidase or 1, 2 α-L- fucosidase. α-L- fucosidase catalyses the reaction between α-L- fucoside and water resulting in the formation of L- fucose as an end product. They belong to the family of glycoside hydrolases that hydrolyse O and S glycosyl compounds. They are also known as α-L- fucoside fucohydrolase. It is involved in n-glycan and glycan structure degradation. In the presence of 4- methylumbelliferyl α-L- fucoside, 1, 2-α-L- fucosidase catalyses the release of 4-methylumbelliferone as a byproduct of the reaction and this can be used in its fluorimetric assay. The substrate used in the colorimetric assay of α- L- fucosidase was para- nitrophenyl- α- D- fucoside.

10. Results and discussion of the proteomic analysis of *Bacteroides fragilis* NCTC 9343 and *Enterobacter cancerogenus* ATCC 35316

Enterobacter cancerogenus and *Bacteroides fragilis* are some of the predominant flora of the gut and are capable of acting as opportunistic pathogens by causing gut associated infections. *B. fragilis* originate in the endogenous flora of the mucous membranes and are capable of causing a number of infections including intra abdominal abscesses, soft skin and tissue infections, bacteraemia, gangrene, necrotizing cellulitis and so on. They are also capable of causing severe secondary infections that include central nervous system infections, pleuropulmonary infections and septic arthritis (Pumbwe *et al*., 2006). Since *Bacteroides* spp. are strictly anaerobic in nature, it hard to detect them during specimen culture and analysis. *E. cancerogenus* has been known to occur as a secondary wound infection causing bacteria. They have been isolated from human faeces, found to be associated with the human gut and form a part of the normal human gut microbiota.

A number of studies have been performed on the differential expression of proteins in *B. fragilis* and *Enterobacter* sp. in response to oxidative stress, antibiotic and so on. In order to determine the proteins that may be differentially expressed in metronidazole-resistant strains of *B. fragilis*, mutants lacking the flavodoxin and pyruvate-ferredoxin oxidoreductase genes were generated and studied (Diniz *et al.*, 2004). Other proteomic studies in *B. fragilis* were performed to determine the oxidative stress response (Rocha *et al.*, 2003). A top down proteomics approach has been used to study the differential expression of heat resistant strains (Williams *et al.*, 2005) and osmotic stress response (Riedel & Lehner, 2007) in *E. sakazakii* and selenite resistance in *E. cloacae* (Barasa, 2008). Since *E. cancerogenus* infections are not as commonly occurring as *B. fragilis* infections, no proteomics experiments have been performed to determine its virulence factors.

10.1 Growth analysis of *E. cancerogenus* and *B. fragilis*

The two strains, *Bacteroides fragilis* NCTC 9343 and *Enterobacter cancerogenus* ATCC 35316, were grown anaerobically at 37°C in an anaerobic chamber. The presence of Gram-negative bacteria was confirmed by Gram staining (Refer to the Figs.12 and 13).

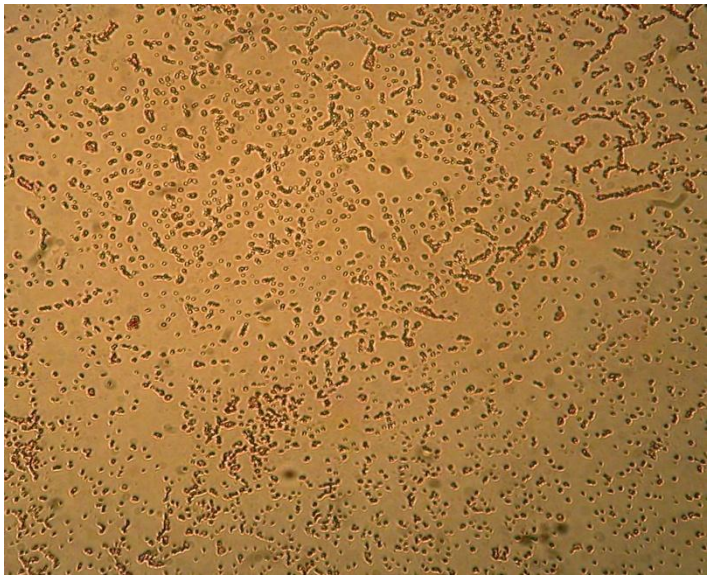


Figure 12: Gram staining results for *E. cancerogenus* showing the presence of small rod shaped bacilli that stained pink.

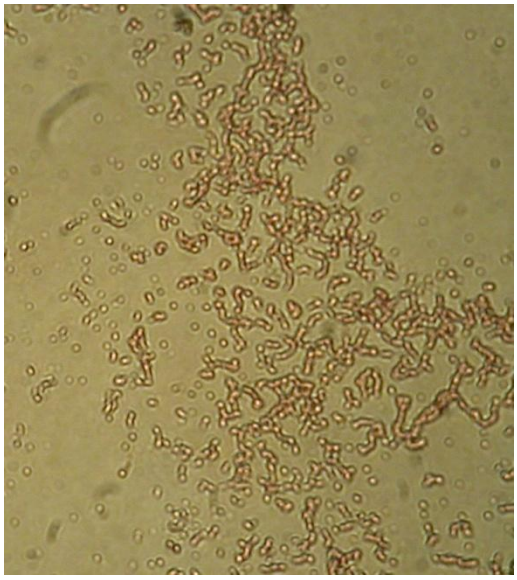


Figure 13: Gram staining results for *B. fragilis* showing the presence of small rod shaped bacilli that stained pink.

Minimal medium was prepared as described in the methodology section and the bacteria were grown in the media. The growth rate of both *B. fragilis* and *E. cancerogenus* was found to be very slow taking about 160 to 170 h to reach an OD of 1.0 at 600 nm. Hence modifications were made in the minimal media where yeast extract was used as a source of vitamin B and the concentration of casitone was increased from 0.2 g to 0.5 g per 100 mL of medium. The growth in this medium (semi- defined media) was found to be much quicker with the culture attaining an overnight stationary phase in case of *E. cancerogenus* (12-16 h) and 40 h for *B. fragilis.* Table 3 showing the growth of *E. cancerogenus* and *B. fragilis* in minimal growth media have been included below. Table 4 included below details the growth of bacteria (*E. cancerogenus* and *B. fragilis*) in semi-defined media containing various sources of carbon.

Time in h	Average absorbance at 600 nm	
	E. cancerogenus	***B. fragilis***
0	0.000	0.000
12	0.008	0.015
24	0.023	0.044
48	0.158	0.102
72	0.278	0.304
96	0.488	0.417
120	0.821	0.632
144	1.025	0.954
168	1.238	1.189
192	1.143	1.178

Table 3: Growth of *E. cancerogenus* and *B. fragilis* in minimal media

Type of carbon source used in the semi-defined media	Average OD at 4 h after inoculation		Average OD at 8 h after inoculation		Average OD at 16 h after inoculation (overnight incubation)	
	E. cancerogenus	***B. fragilis***	***E. cancerogenus***	***B. fragilis***	***E. cancerogenus***	***B. fragilis***
Glucose	0.839±0.0008	0.223±0.0009	1.036± 0.002	0.539±0.0012	1.304± 0.0025	0.728±0.0008
Mucin Type II	0.378±0.001	0.035±0.0008	0.367± 0.156	0.156±0.0008	0.358± 0.002	0.462±0.0017
Mucin Type III	0.355 ±0.0005	0.038±0.0012	0.353± 0.093	0.201±0.0008	0.325± 0.0008	0.512±0.0008
Glucose and Type II mucin (0.5 g/100 mL of glucose and mucin Type II)	0.993±0.0012	0.324±0.0012	1.148 ±0.001	0.498±0.0012	1.316±0.0016	0.819±0.0017
Glucose and Type III media (0.5 g/100 mL of glucose and mucin Type III)	1.090±0.0124	0.287±0.0009	1.220±0.025	0.562±0.0016	1.414 ±0.0017	0.753±0.0016
Type II and III mucin (0.5 g/100 mL of mucin Type II and Type III)	0.956±0.0008	0.045±0.001	1.134±0.002	0.198±0.0008	0.614±0.0012	0.485±0.0012
Glucose and Type II and III mucin (0.5 g/100 mL of glucose, mucin Type II and Type III)	0.068±0.0017	0.015±0.0017	0.881±0.0017	0.367±0.0008	0.821± 0.0012	0.328±0.0008

Table 4: Growth of bacteria in the semi-defined media with different carbon sources

All experiments were performed in duplicate and the values represented in the tables indicate average values of absorbance (600 nm) at the respective time points. All the carbon sources were added at a concentration of 0.5 g per 100 mL of the media. A quantity of 0.5 g of mucin was weighed out from the bottle and added to the medium rather than measuring the concentration of sugar in it.

10.1.1 Growth analysis of *E. cancerogenus*

Media (450 mL) inoculated with a 1 mL volume from 10 mL overnight starter cultures showed the following growth results over a period of 8 h (Fig. 14). The optical density values were measured at 600 nm at various time periods. An average of three results is represented in the graph below for both OD values against time and three different media have been used.

All the growth curve experiments were performed in triplicate and vertical error bars representing the standard deviation from the mean can be observed in the graph.

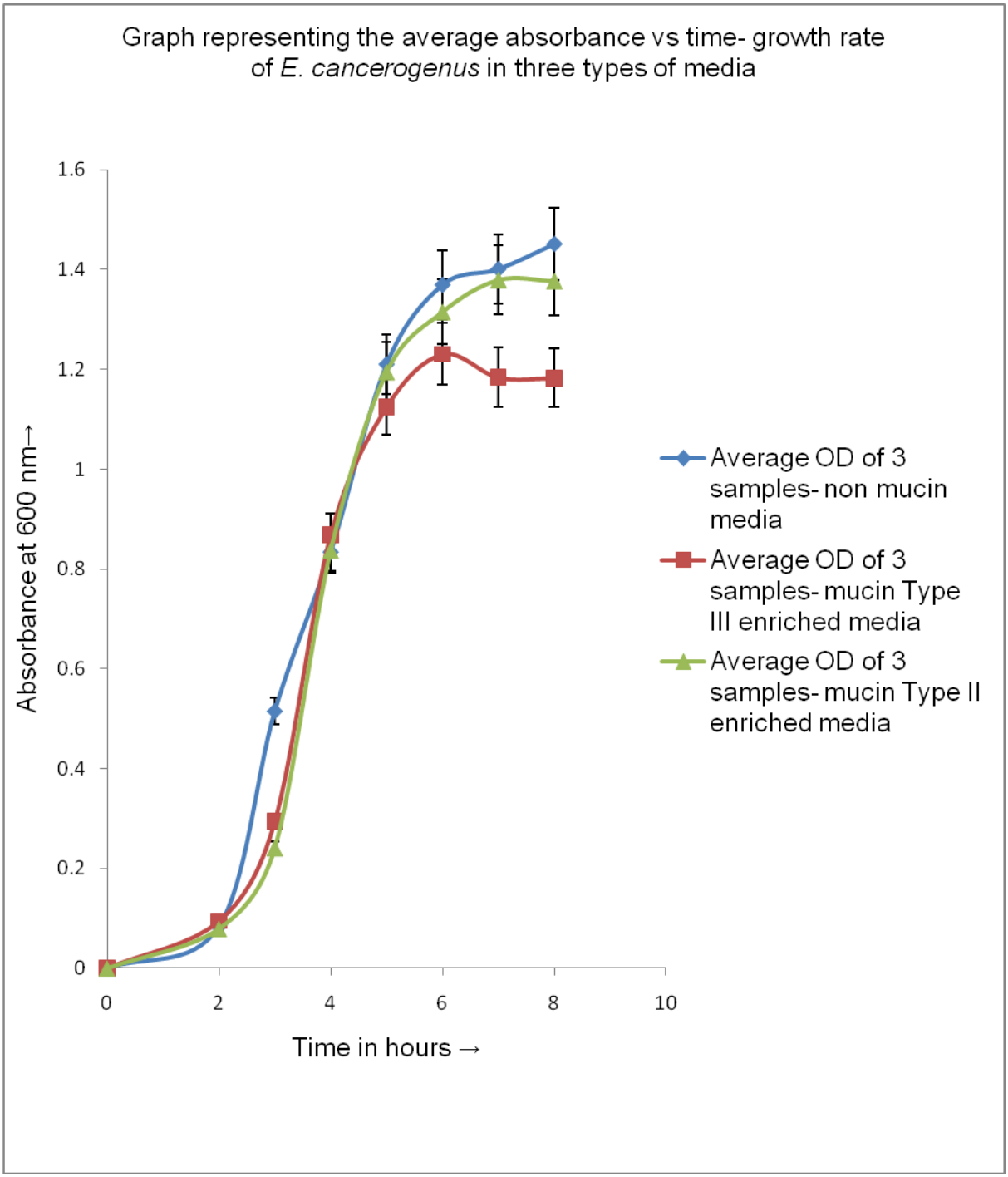


Figure 14: Graph representing the average absorbance vs time- growth rate of *E. cancerogenus* in three types of media

10.1.2 Growth analysis of *B. fragilis*

Media (450 mL) inoculated with a 1 mL volume from 10 mL overnight starter cultures showed the following growth results over a period of 48 h (Fig. 15). The optical density values were measured at 600 nm at various time periods. An average of three results is represented in the graph below for both OD values and colonies forming units against time and three different media have been used.

All the growth curve experiments were performed in triplicate and vertical error bars representing the standard deviation from the mean can be observed in the graph.

The average absorbances for the growth of *E. cancerogenus* and *B. fragilis* in semi-defined media have been included in the Appendix section E in the following tables. Tables 5, 6 and 7 included in the Appendix E show the growth of *E. cancerogenus* in non mucin, mucin Type II and Type III enriched media, respectively. Tables 8, 9 and 10 included in the Appendix E show the data for the growth of *B. fragilis* in non mucin, mucin Type II and Type III enriched media, respectively.

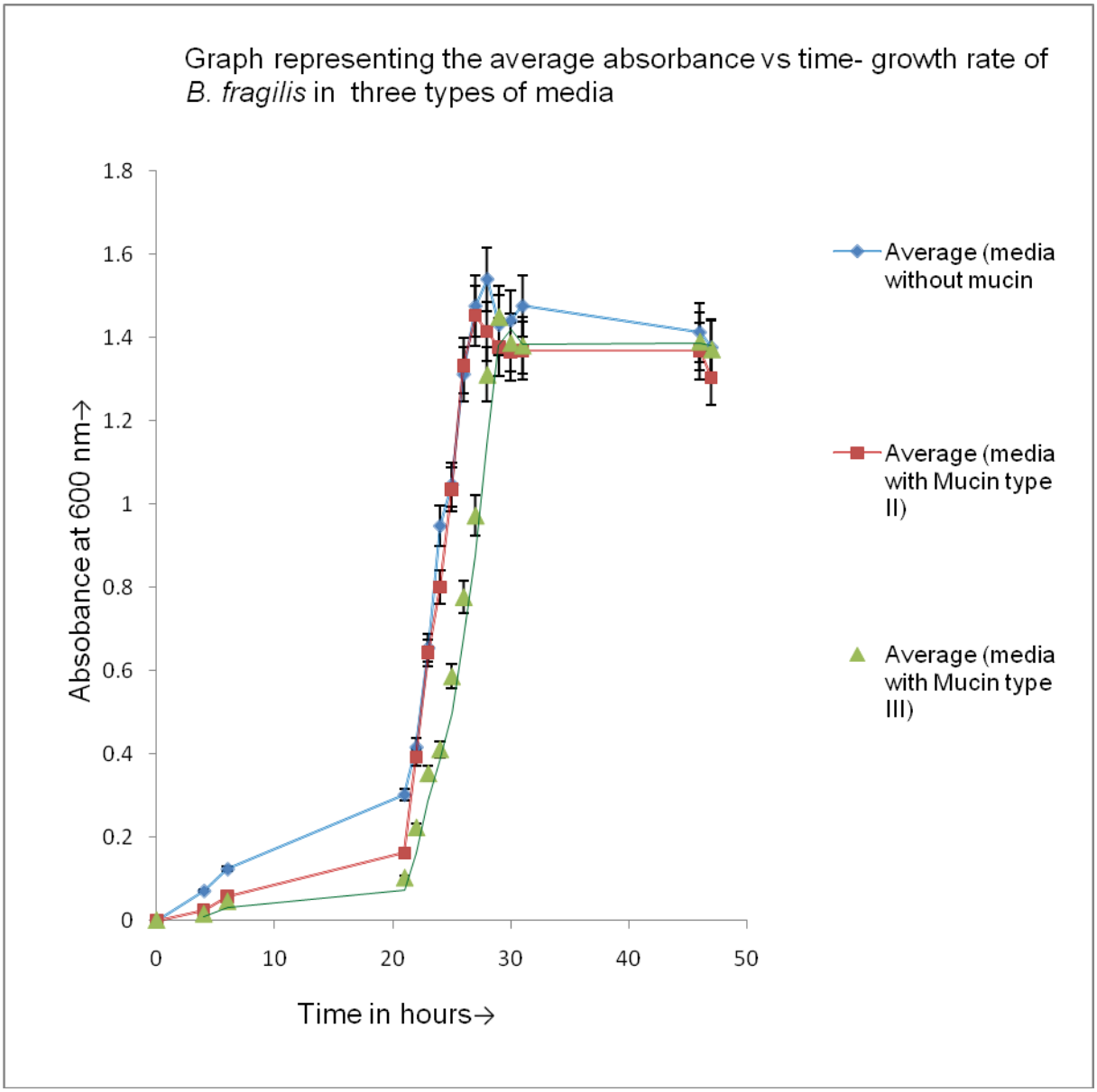


Figure 15: Graph representing the average absorbance vs time- growth rate of *B. fragilis* in three types of media

The harvesting time point for the cultures of *E. cancerogenus* and *B. fragilis* were determined based on halving the maximum OD value achieved during its growth in the semi-defined media as 0.6 and 0.7, respectively. This would also ensure that the cells were in their mid log phase.

10.2 Growth of *E. cancerogenus* ATCC 35316 and *B. fragilis* NCTC 9343 in semi-defined media

Both bacteria were grown in three types of semi-defined media which included media without mucin, media enriched with porcine intestinal mucin Type II and media enriched with porcine intestinal mucin Type III. Even though initial attempts were aimed at growing the bacteria in minimal media with and without mucin, it was found that the extremely slow rate of growth of bacteria hindered the progress of experiments and hence the medium was slightly modified by the addition of yeast extract, replacing the vitamin solution and increasing the concentration of casitone from 0.2 to 0.5 g to enable quicker growth of the bacteria.

Growth of the bacteria was studied in the logarithmic phase rather than the stationary phase because most of the virulence factors and other important proteins are normally expressed in the logarithmic phase of growth (Ernst *et al*., 1990).

Several carbon sources were used to study the growth of *E. cancerogenus* and *B. fragilis* and it was found that the semi-defined medium containing glucose on its own (or in combination with mucin) supported efficient growth. In a semi-defined medium containing only mucin as a source of carbon, the growth rate was found to be low, with the optical density not increasing beyond 1.0, when compared to a semi-defined medium containing glucose and mucin. This may be due to the general tendency of bacteria to use glucose as the first source of carbon and other carbohydrate sources are utilised only when glucose is completely depleted in the media (Studier, 2005).

10.3 Mucin used in the study

The two types of mucin used were porcine gastric/intestinal mucin Type II and III. The main difference between the Type II and Type III medium is that mucin Type III is a partially purified powder that has the presence of 0.5 to 1.5% bound sialic acid. The sugar and amino acid composition of mucin has not been determined by the Sigma-Aldrich company from where it was purchased. The Type II mucin is a crude extract of mucin glycoproteins from the hog stomach using the enzyme pepsin and it has about 1% of bound sialic acid. Since the study aimed at analysing the differential protein expression in opportunistic pathogens when they come in contact with the

first line of cell defence, mucin, the closest possible system to the human body was necessary. Previous studies show that the core structures found in porcine gastric mucins are the same as that in human gastric mucins (Karlsson *et al.*, 1997). One of the most easily accessible types of mucin was the porcine gastric mucin without the involvement of any ethical issues thereby encouraging the incorporation of mucin as a potential source of carbohydrates in the semi-defined medium.

Out of several mucin types used in a study, ovomucoid, porcine gastric mucin and beef submaxillary mucin were found exhibit high similarity to human colonic mucin (Salyers *et al*., 1977). The variations mainly occur in a relatively small number of basic structures. Some of these variations include the presence of acidic structures derived from neutral oligosaccharides in human colonic mucin, variations in the degree of sialylation, the biantennary structures of mucin oligosaccharides and also the branched chain olicosaccharides which appear to be linear chain derivatives of the same core structure (Podolsky, 1985). High levels of sialylation have been observed in both human and porcine colonic mucin. The hydroxyl groups present at the end of the sialic acids that contain acetyl esters inhibit bacterial degradation of mucin (Corfield *et al*., 1992). Disulphide rich domains have been idenitified in both human intestinal and porcine submaxillary mucins but there may be a low degree of sequence identity (PerezVilar *et al*., 1996).

It is possible that the activity of *B. fragilis* on colonic mucin may be different to that of gastric mucin since gastric mucin has been used in our studies due to the unavailability of colonic mucin.

My results show the expression of a number of proteins related to oxidative stress and heat shock which indicate the presence of stress during growth. Previous proteomics studies in *E. faecalis* have revealed the expression of a number of stress associated proteins in exponentially grown cells, viable but non culturable cells and proteins expressed during starvation conditions (Heim *et al*., 2002). These stress proteins are generally expressed in response to bacterial growth and adaptability to various environmental conditions.

Proteome profiles obtained from the growth of bacteria in mucin enriched media suggest that the differentially expressed proteins may be associated with colonisation where bacteria adapt themselves to bind to the mucous layers of the

gut. Bacterial proteins associated with the ability to induce an immune response within the host could act as potential virulence factors and eventually result in infection. In the proteomics studies, first dimension separation of proteins was limited to a pH range of 4-7 rather than a variety of different pH gradients. Performing experiments with different second dimension separations may also provide a better insight into understanding differential protein expression. There are a large number of bacteria that occur commonly on mucous membranes forming a part of the normal gut flora by adhering and colonising mucin. But certain environmental adaptations within them can result in their ability to produce enzymes and compounds that lead to virulence.

10.4 Optimisation of two dimensional polyacrylamide gel electrophoresis (2DE-PAGE) gels

10.4.1 Comparison of staining methods

An attempt was used to compare the Coomassie blue staining with fluorescent staining, but due to the inconvenience of the requirement to irradiate with UV light when excising protein spots and poorer resolution, the Coomassie blue staining technique was chosen to analyse the spots of interest. On comparison of the gels represented in Figs 16.1 and 16.2 it is evident that the spots have been expressed with better clarity and resolution when colloidal Coomassie blue stain was used. From the figure, it was evident that fluorescent staining failed to achieve the required sensitivity levels and seemed inconsistent when compared to Coomassie blue staining. The process of cutting out spots of interest would be a more tedious process if fluorescent staining was used since this would involve working in the presence of UV light. The gel looks smudgy and the presence of a dark background may obscure the number of low abundance proteins observed. The improper background staining may be attributed to the trapping of the fluorescent dye in detergent micelles and this has been observed in previous studies (Malone *et al*., 2001).

Figure 16: Figure showing the difference between fluorescent staining and Coomassie blue staining.

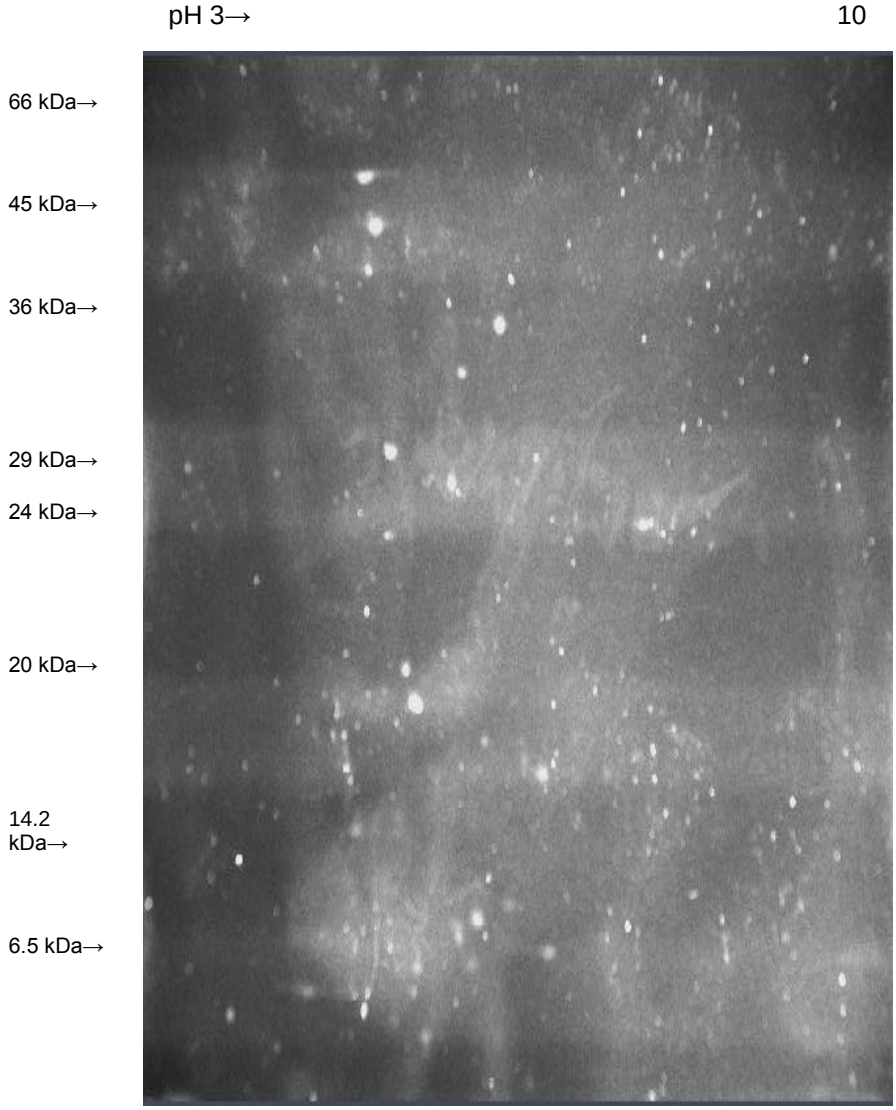


Figure 16.1: Proteome map of *E. cancerogenus* grown in semi-defined media without mucin. Separation was achieved using pH range 3-10 IEF strips. Gel was stained using fluorescent staining.

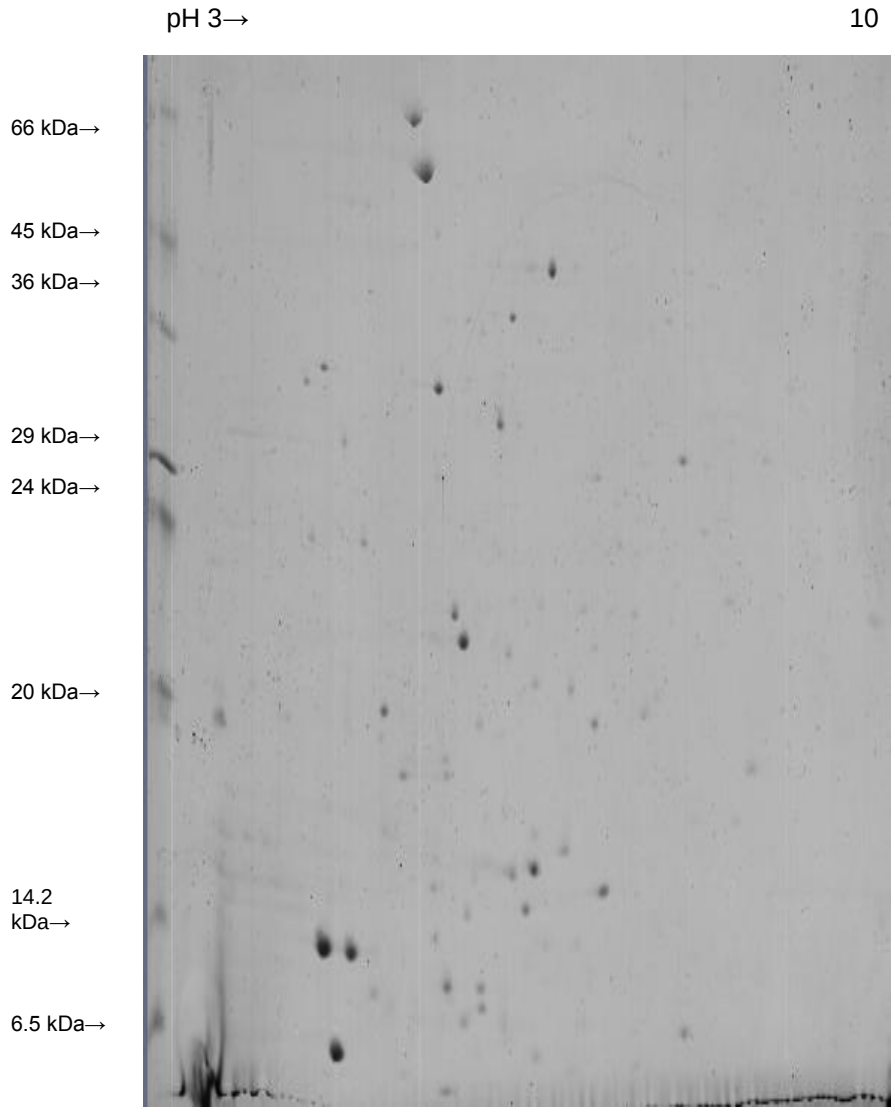


Figure 16.2: Proteome map of *E. cancerogenus* grown in semi-defined media without mucin. Separation was achieved using pH range 3-10 IEF strips Gel was stained using colloidal Coomassie blue staining.

10.4.2 Comparison of pH ranges of IEF strips

The proteins from the culture supernatant were initially analysed using pH range 3-10 IEF strips. However the location of proteins seemed to be predominantly within the 4-7 pH range (Figures 17.1, 17.2 and 17.3) and hence the IPG strip pH range was changed to 4-7.

Figure 17: Isolation of E*. cancerogenus* proteins from 450 mL mid log phase cultures at an OD of 0.6 separated using pH range of 3-10 IEF strips

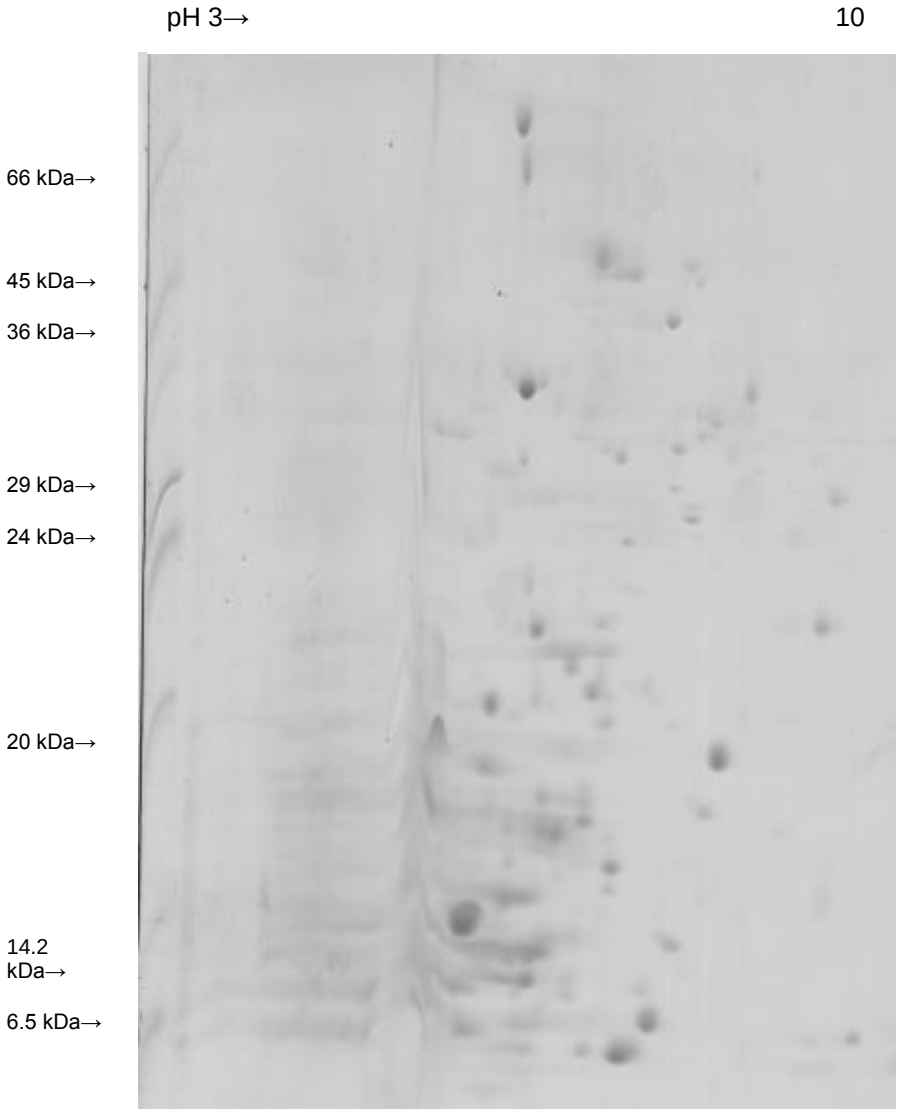


Figure 17.1: Proteome map of *E. cancerogenus* grown in semi-defined media without mucin. Separation was achived using pH range 3-10 IEF strips.

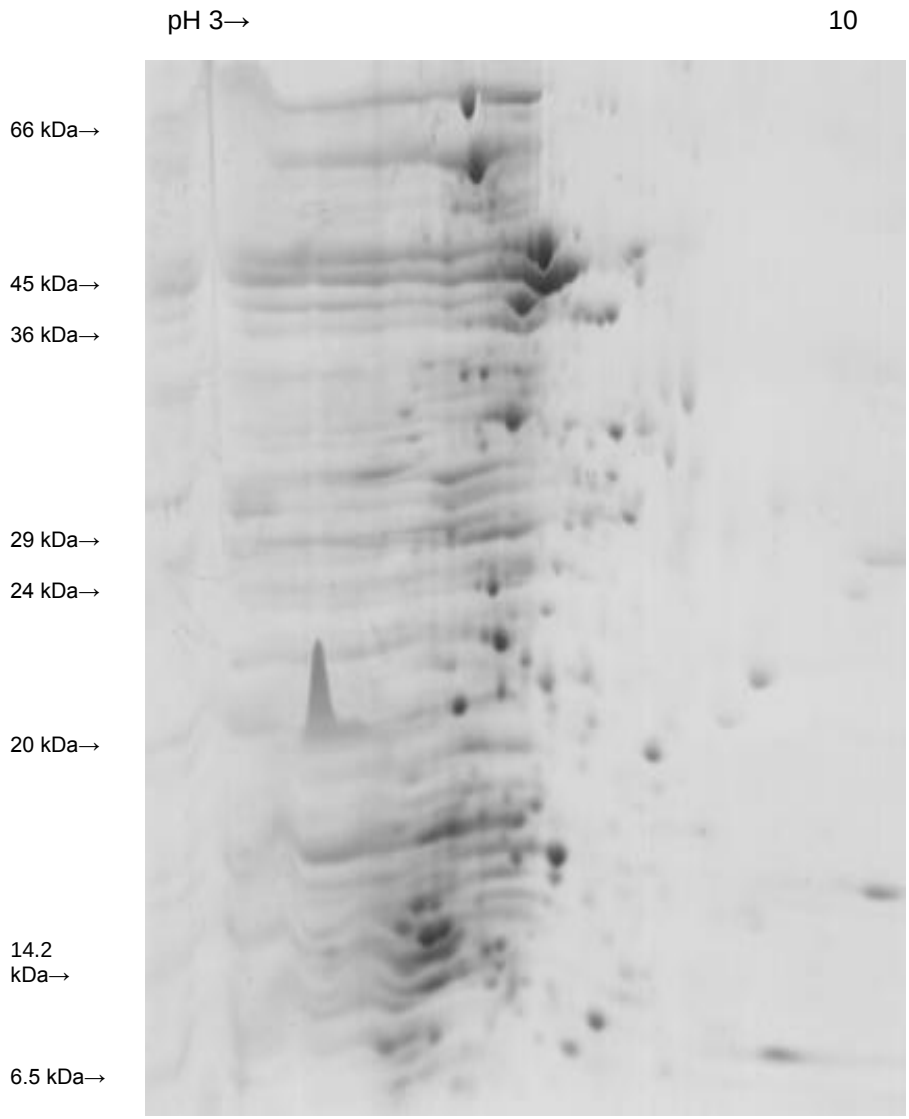


Figure 17.2: Proteome map of *E. cancerogenus* grown in semi-defined media enriched with mucin Type II. Separation was achieved using pH 3-10 IEF strips.

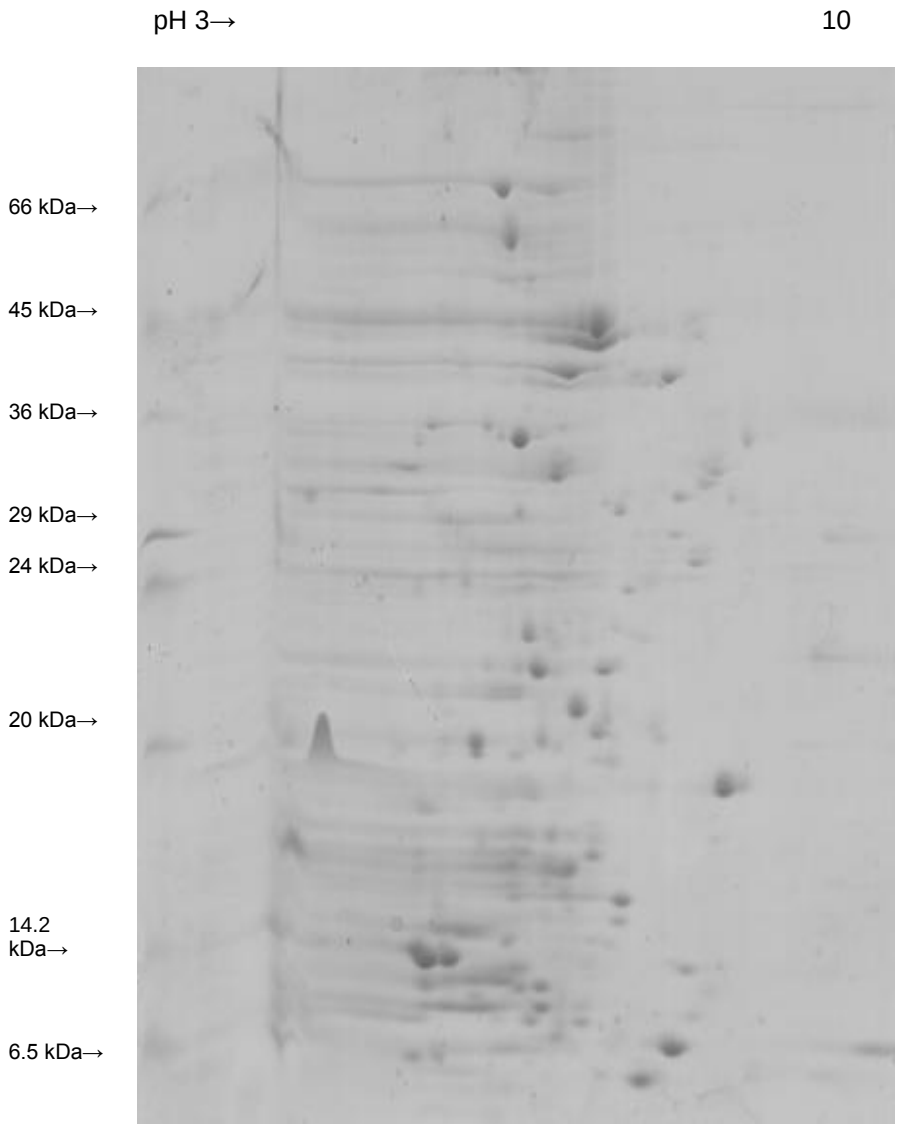


Figure 17.3: Proteome map of *E. cancerogenus* grown in semi-defined media enriched with mucin Type III. Separation was achieved using pH 3-10 IEF strips.

10.4.3 Comparison of protein loadings

The expression of proteins was studied in the mid logarithmic phase of growth of the culture and several optimisation experiments were performed to determine the appropriate volume of the culture that could provide a good separation and resolution of spots. Initial experiments were performed using 450 mL of the log phase cultures and were later reduced to a volume of 100 mL in order to obtain the best possible resolution (Figs. 18.1 and 18.2).

Figure 18. Proteins isolated from 100 and 450 mL of *B.fragilis* and *E. cancerogenus* grown in semi-defined media separated by pH 4-7 IEF strips

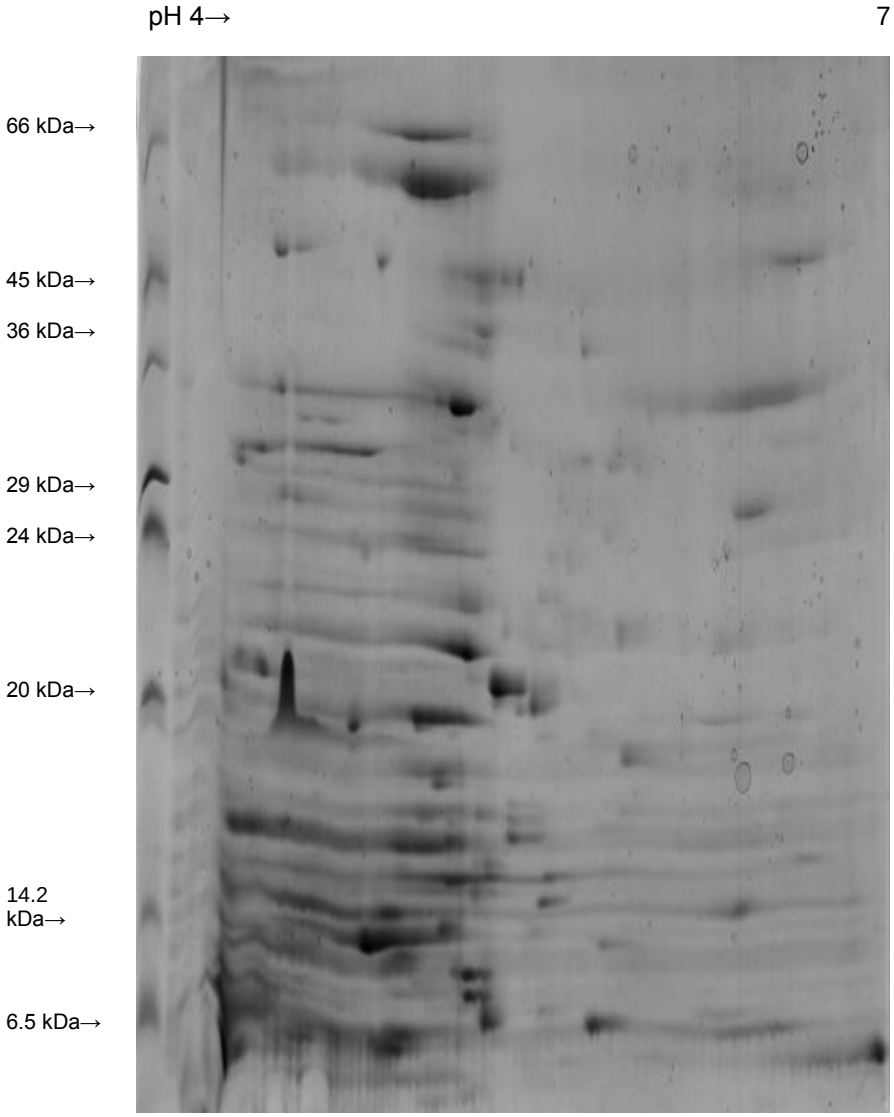


Figure 18.1: Proteins isolated from 450 mL of *E. cancerogenus* grown in semi-defined media without mucin separated by pH 4-7 IEF strips.

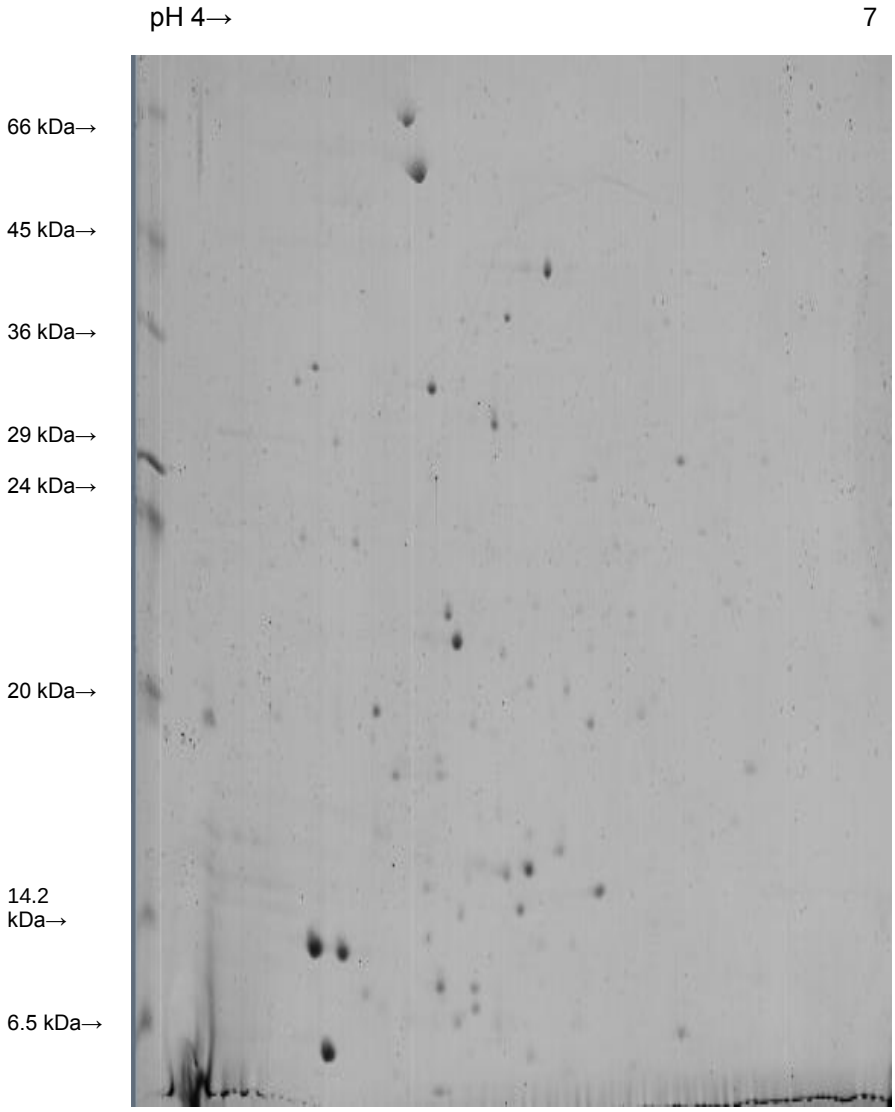


Figure 18.2: Proteins isolated from 100 mL of E.*cancerogenus* grown in semi-defined media without mucin separated by pH 4-7 IEF strips.

10.4.4 Comparison of polyacylamide gel concentrations

The protein expression patterns were found to be more distinct and better resolved in 12 % (w/v) rather than 14% (w/v) gels and hence 12 % (w/v) gels were used during the study (Figures 19.1 and 19.2).

Figure 19: Isolation of E*. cancerogenus* proteins from 100 mL log phase cultures at an OD of 0.6 separated using pH range of 4-7 IEF strips

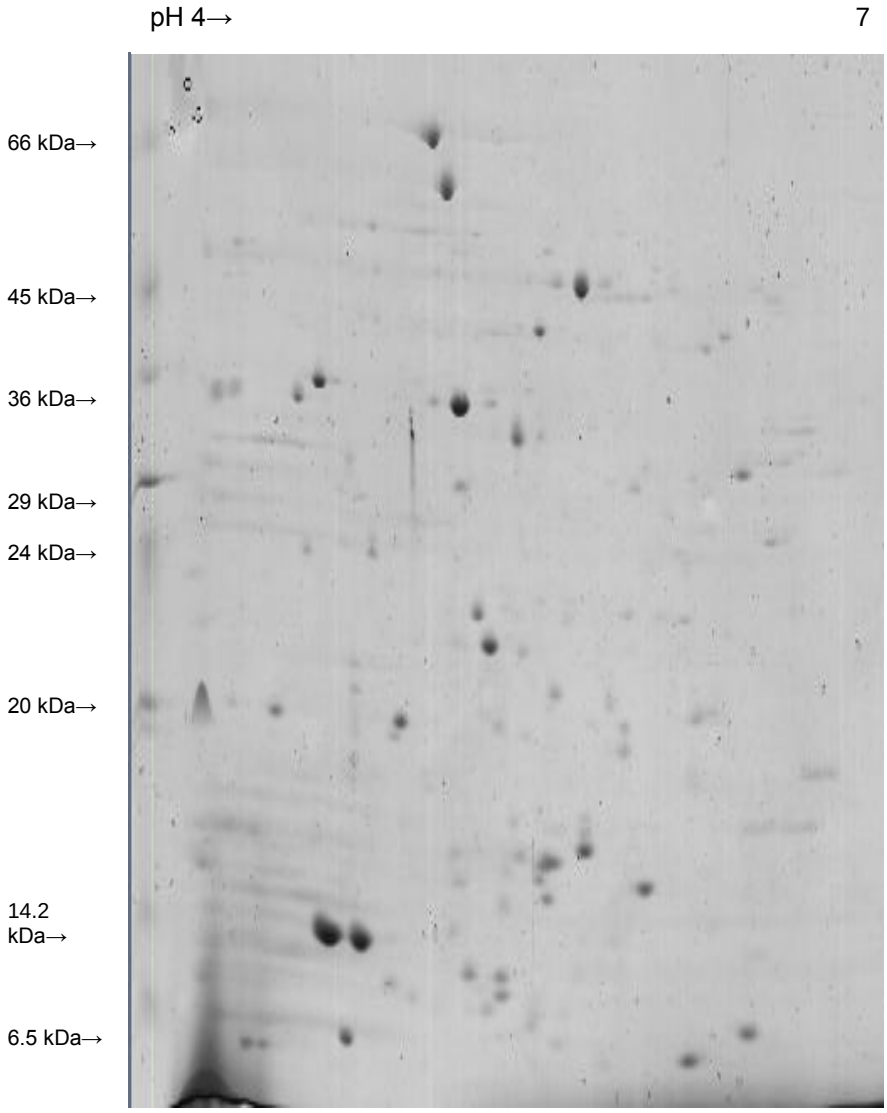


Figure 19.1: Proteome map of *E. cancerogenus* grown in semi-defined media without mucin. Separation was achieved using a 12% (w/v) gel

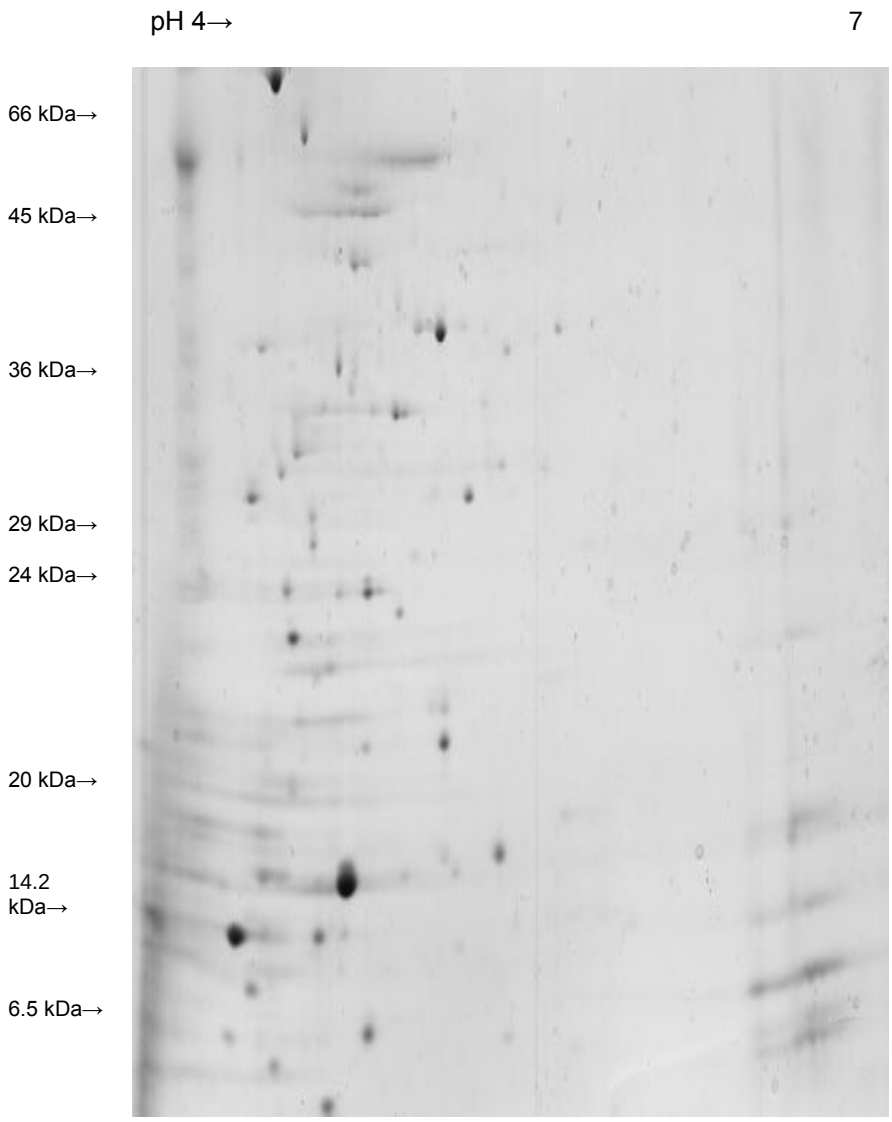


Figure 19.2: Proteome map of *E. cancerogenus* grown in semi-defined media without mucin. Separation was achieved using a 14% (w/v) gel

10.5 Two dimensional gel electrophoresis

One of the initial problems was the culture volume required to obtain a protein expression profile with good resolution and minimising horizontal streaks. Hence different volumes of cultures were used to obtain expression patterns in the log phase and a 100 mL culture volume at an OD_{600nm} of 0.6 or 0.7 in *E. cancerogenus* or *B. fragilis*, respectively, was found to be the most appropriate. Based on the results obtained, 100 mL culture volumes were used in the study (Refer to Figs. 18.1, 18.2 with different volumes in results section).

Sample preparation is one of the most important parts of the two dimensional gel electrophoresis (2DE) technique since even small variations in methodology could

result in massive changes in protein expression profiles. Reproducibility of gels was another major issue faced in the 2DE method. Care was taken to ensure that proteins studied from different conditions were extracted and analysed in a similar fashion aiming at minimising variation as much as possible (Choe & Lee, 2003). At least six gels were generated from each condition in order to confirm that they were reproducible. During Mann-Whitney U analysis using the PDQuest and Redfin softwares, spots that were not up- or down-regulated in all the replicate gels compared to all the replicated control gels were excluded. During Student t-test analysis using the PDQuest, Redfin and SameSpots software, an average expression in each condition was considered rather than the expression levels in each replicate gel.

Initial experiments resulted in a high amount of horizontal and vertical streaks on the gels with poor resolution. One of the main reasons for this was the incomplete resolubilisation of pellets following their treatment with the 2D Clean up kit. The protein pellets were resolubilised in the rehydration buffer by ultra-sonication on ice for a few seconds and the insoluble pellets were removed by centrifugation for 1 min. This helped to reduce protein overload on to the IPG strips and also resulted in efficient separation of proteins.

The composition of the solutions used in the 2D Clean up kit are not freely available. Previous studies have shown that they help to remove contaminating polysaccharides from the protein samples (Zhang, 2006) (Zhang *et al*., 2007).The 2D Clean up kit consists of the precipitant, co-precipitant, wash buffer and wash additive solutions whose recipe has been patented. Alternatively, the acetone precipitation method could be used to prepare protein samples from *B. fragilis* cultures where acetone is the main compound precipitating out the proteins present in the cell free extract.

Similar studies were performed to determine the pH range for the IPG strips used in the first dimension electrophoresis. The pH range of 3-10 was used initially but since most of the proteome expression was concentrated around the acidic pH range of 4-7, further experiments were performed using IPG strips of pH 4-7 (refer to Figs. 17).

An attempt was used to compare the Coomassie blue staining with fluorescent staining but due to the inconvenience of cutting out spots and improper resolution,

the Coomassie blue staining technique was chosen to analyse the spots of interest. The inconvenience of cutting out spots was that UV light was required to locate the exact spot boundaries and this would make the cutting process a tedious task and visualisation in visible light impossible. From Figs. 16.1 and 16.2, it is evident that the spots have been expressed with better clarity and resolution when colloidal Coomassie blue stain was used. The increased convenience of using colloidal Coomassie blue staining method when compared to fluorescent staining supported its use in the study.

Since reproducibility and uniform dimensions of scan area on gels were some of the most important factors to be considered, the scanning of all the gels were performed together at the same time ensuring that the scan area was exactly the same for all of them. Further adjustments were possible using the PDQuest, Redfin and Samespots software tools where flipping of gels, cropping and background subtractions were carried out to optimise and compensate for differences in intensities of staining and other experimental variations. The spots on the gels were reviewed using the spot review tool and the consensus tool was used to edit any errors that may have occurred during spot analysis.

10.6 Separation of proteins from *E. cancerogenus* and *B. fragilis* grown in semi-defined media with and without mucin using 2DE gels.

10.6.1 Sets of replicate 2DE gels of proteins from *E. cancerogenus*

Multiple sets of gels of biological replicates of proteins from *E. cancerogenus* grown in media without mucin, media with mucin Type II and media with mucin Type III (IEF strip pH range 4-7, processing volume: 100 mL, harvesting OD at 600 nm: 0.6) (Figs. 20-22).

Figure 20: Proteome maps of *E. cancerogenus* grown in semi-defined media without mucin (control media; using pH 4-7 IEF strips).

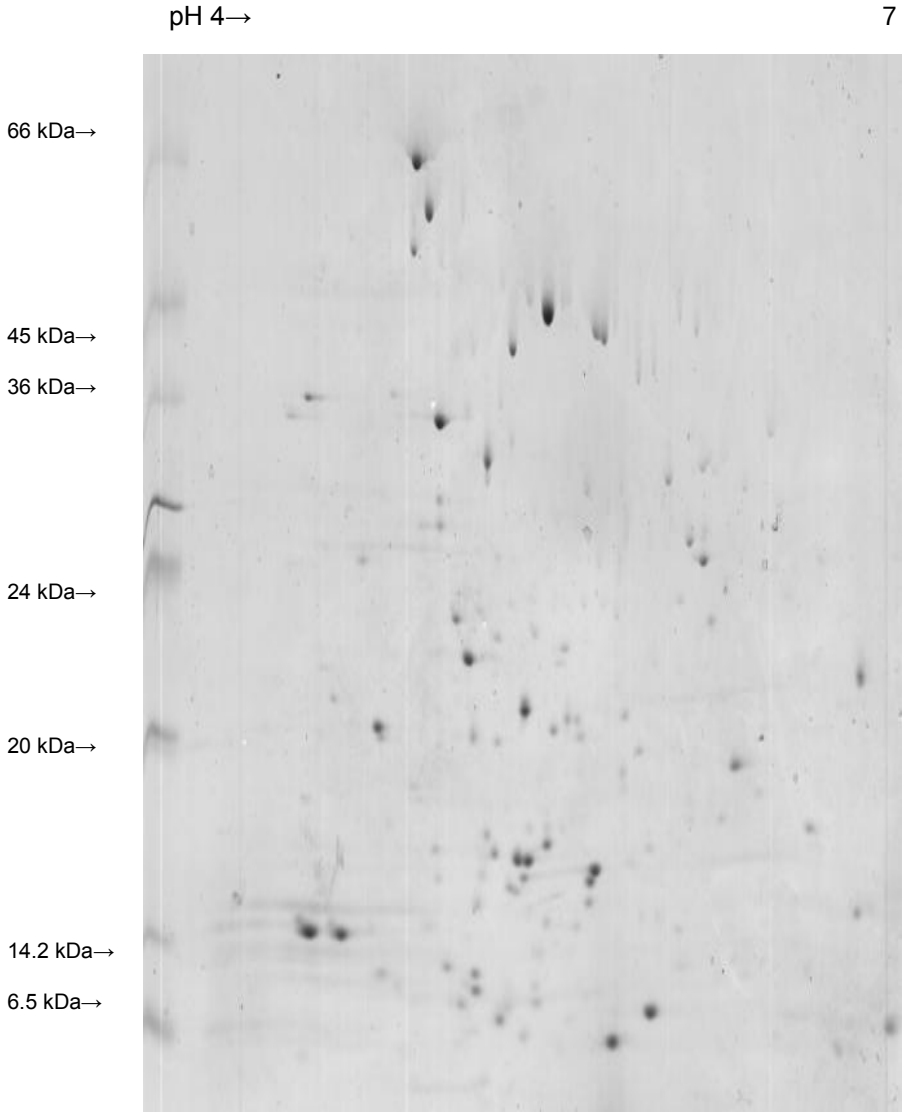


Figure 20.1: Proteome map no.1 of *E. cancerogenus* grown in semi-defined media without mucin

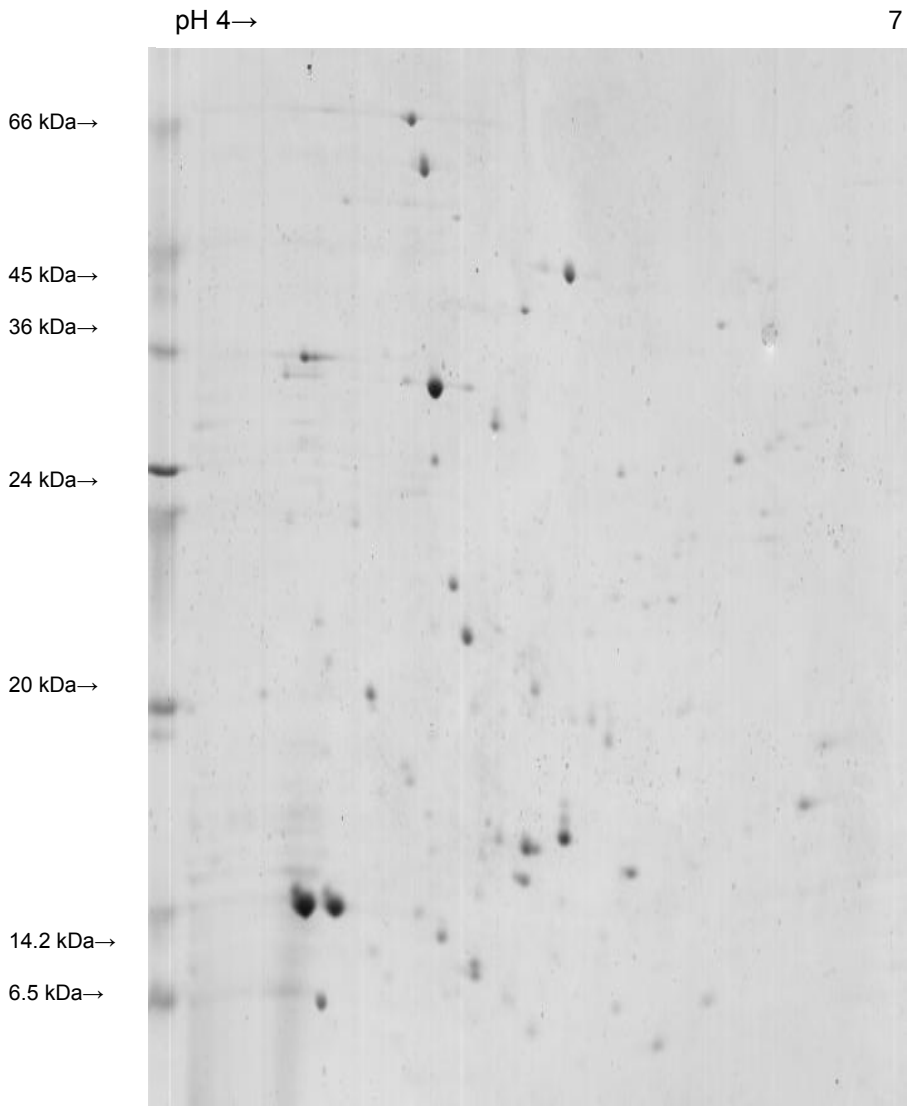


Figure 20.2: Proteome map no.2 of *E. cancerogenus* grown in semi-defined media without mucin.

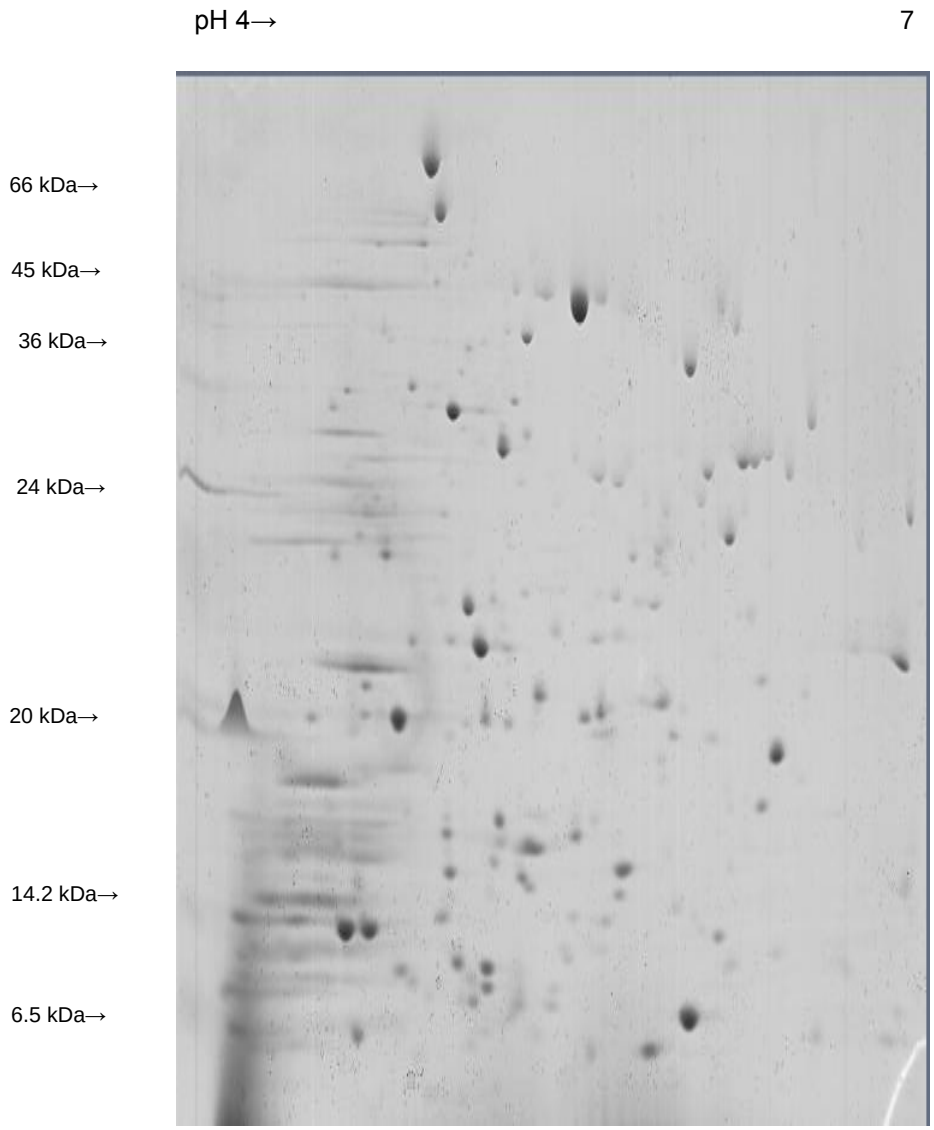


Figure 20.3: Proteome map no.3 of *E. cancerogenus* grown in semi-defined media without mucin.

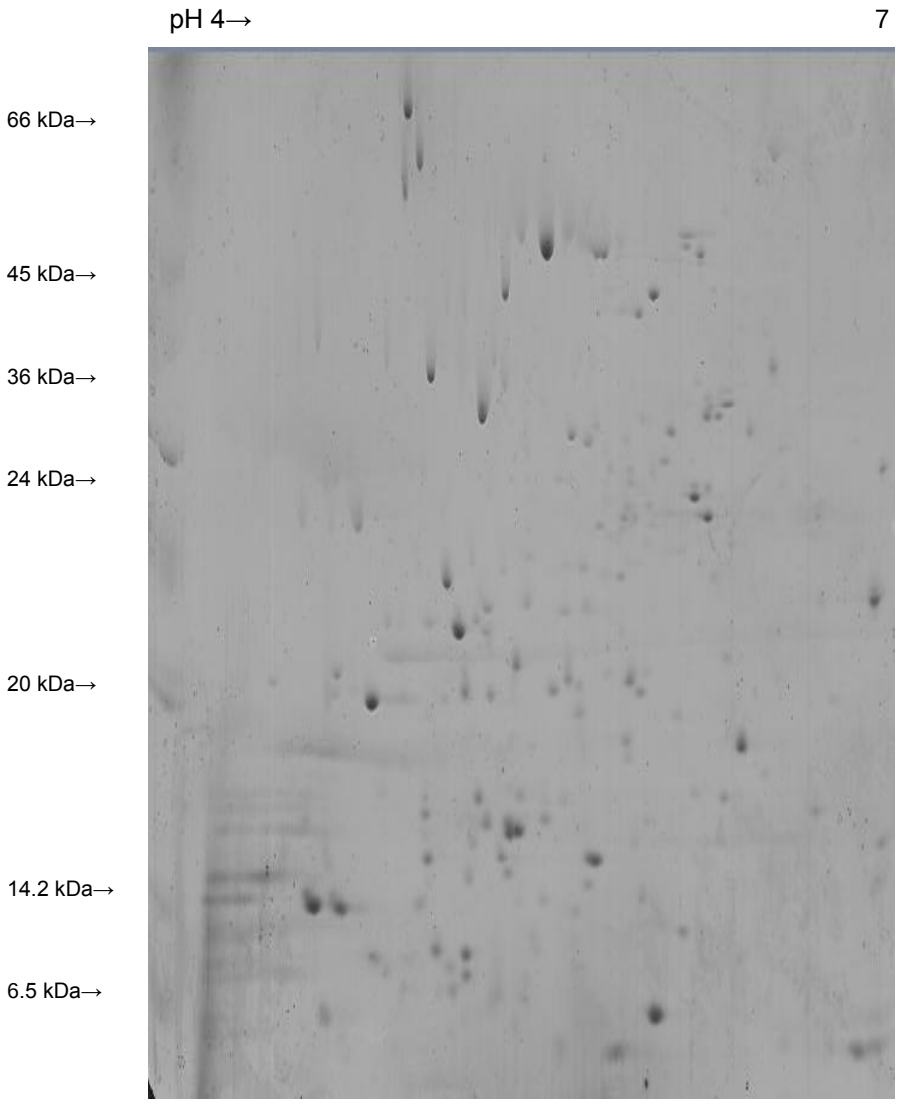


Figure 20.4: Proteome map no.4 of *E. cancerogenus* grown in semi-defined media without mucin.

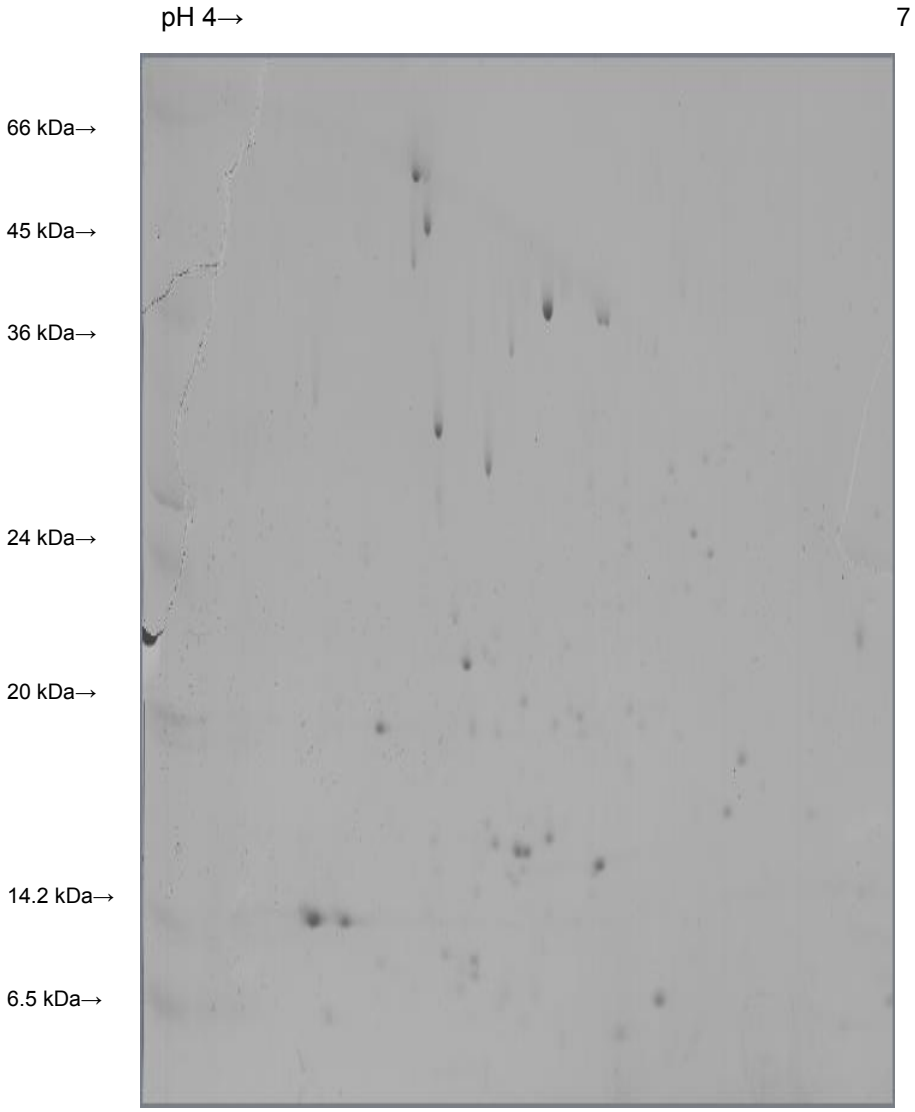


Figure 20.5: Proteome map no.5 of *E. cancerogenus* grown in semi-defined media without mucin.

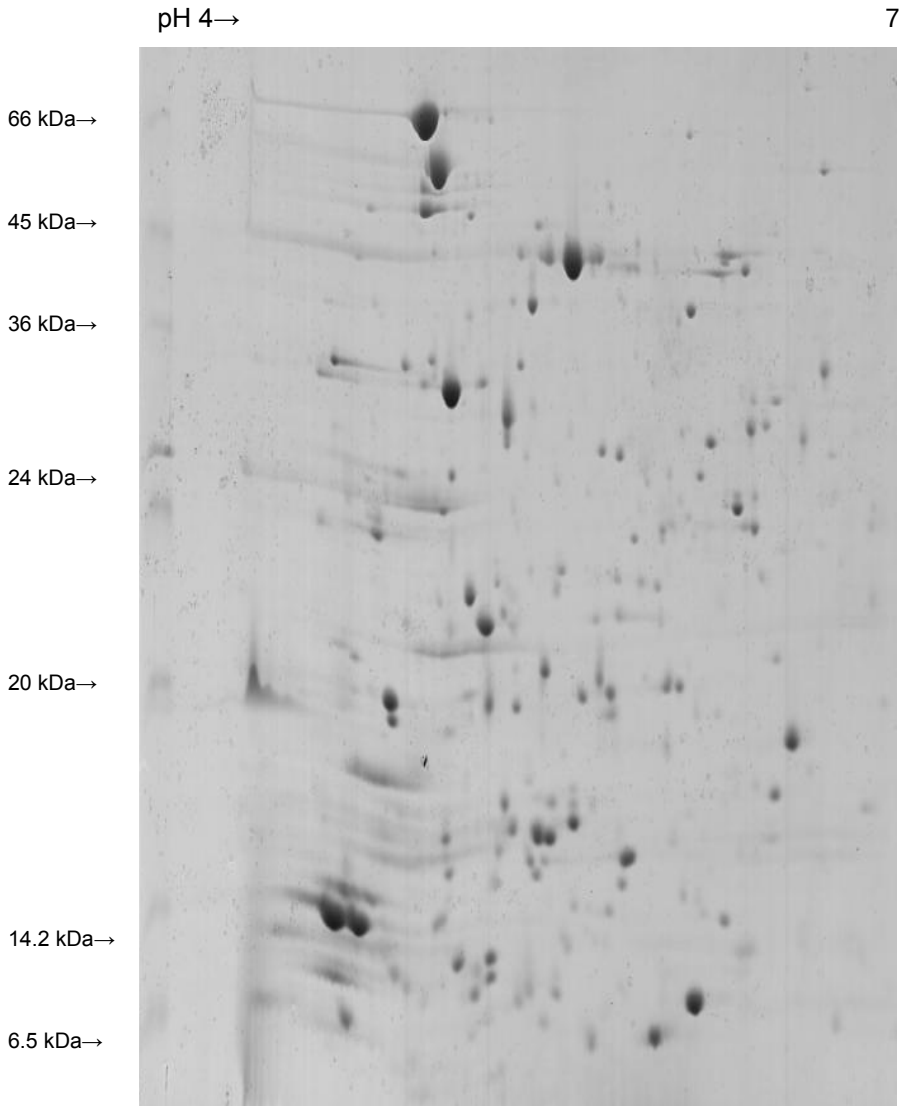


Figure 20.6: Proteome map no.6 of *E. cancerogenus* grown in semi-defined media without mucin.

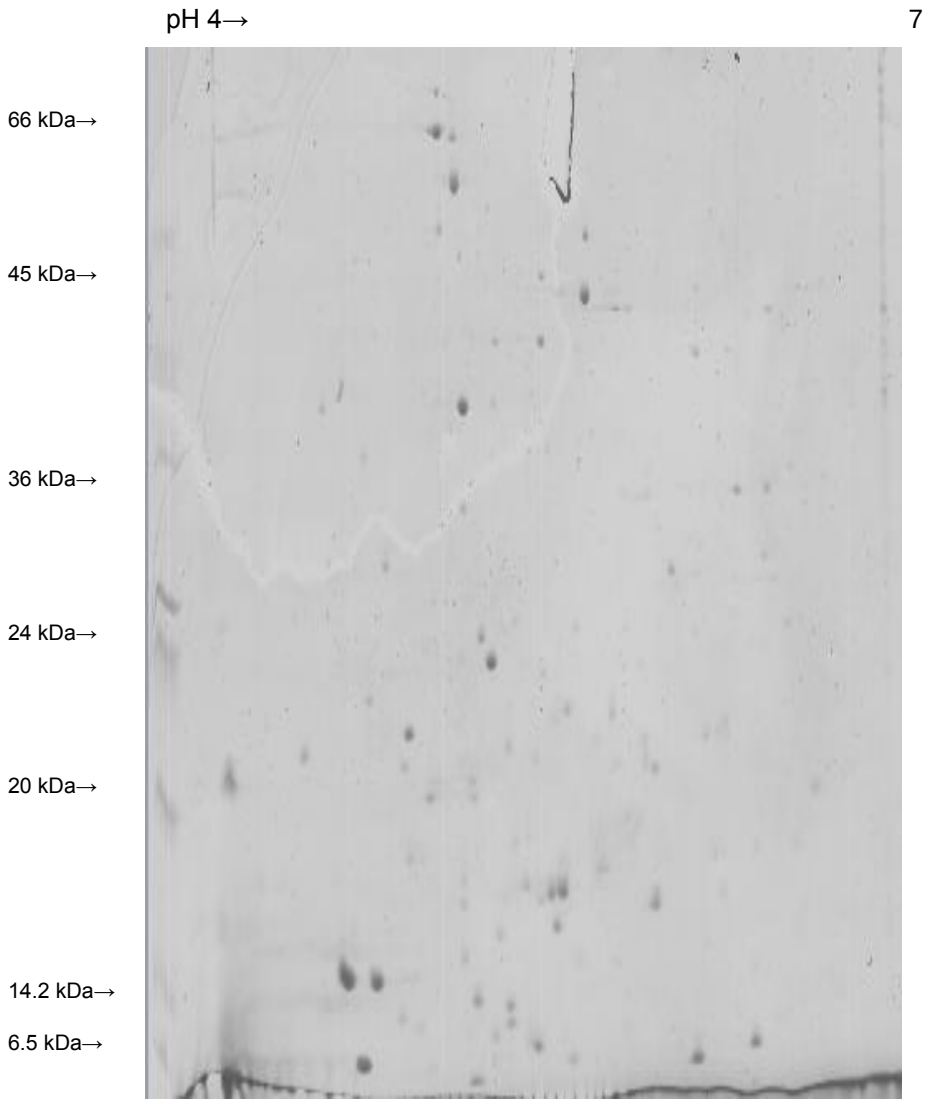


Figure 20.7: Proteome map no.7 of *E. cancerogenus* grown in semi-defined media without mucin.

Figure 21: Proteome maps of *E. cancerogenus* grown in semi-defined media enriched with mucin Type II (using pH 4-7 IEF strips)

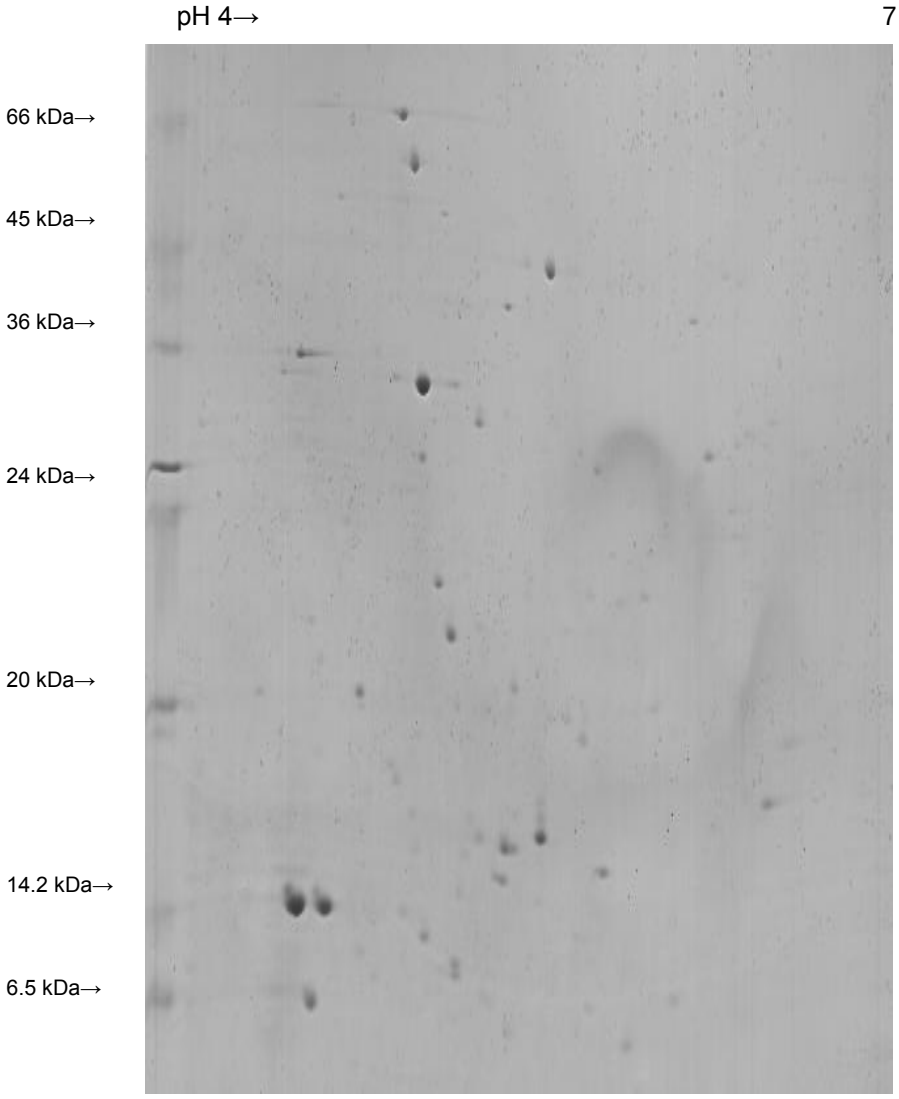


Figure 21.1: Proteome map no.1 of *E. cancerogenus* grown in mucin Type II enriched media

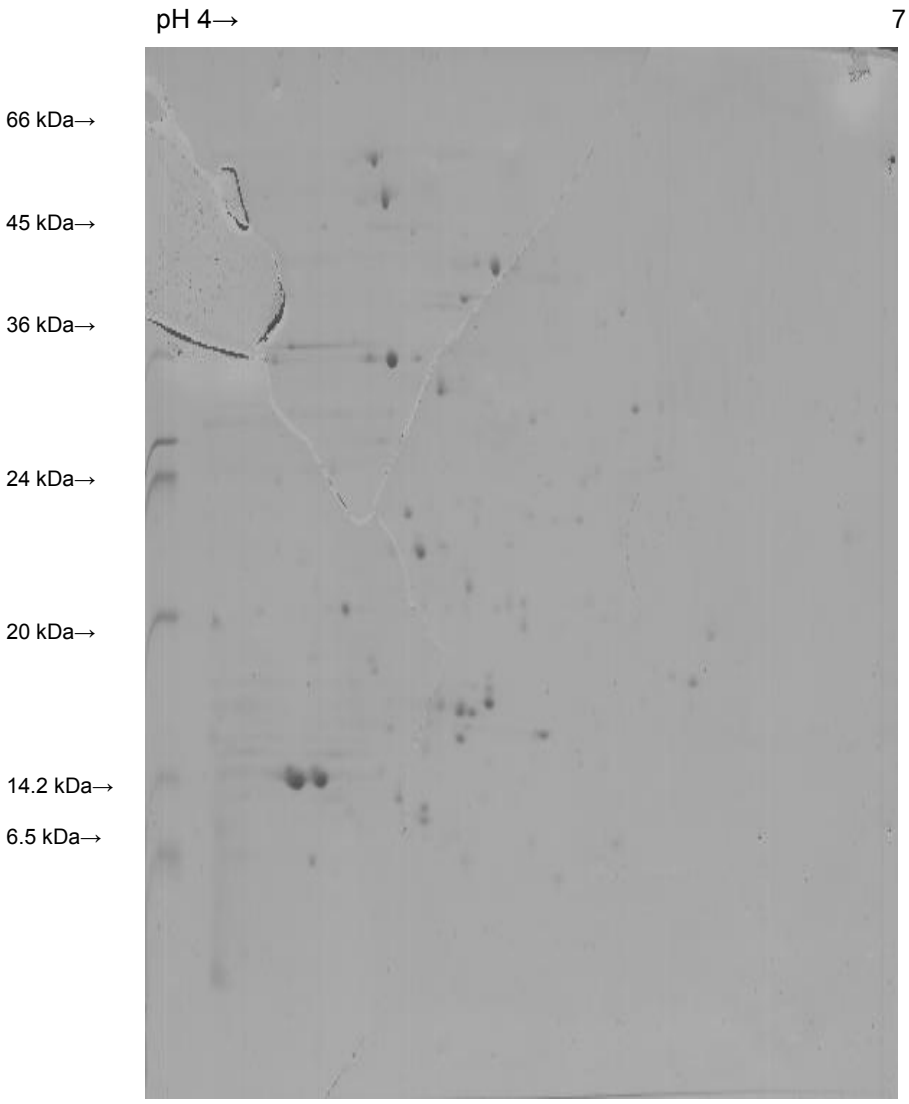


Figure 21.2: Proteome map no.2 of *E. cancerogenus* grown in mucin Type II enriched media

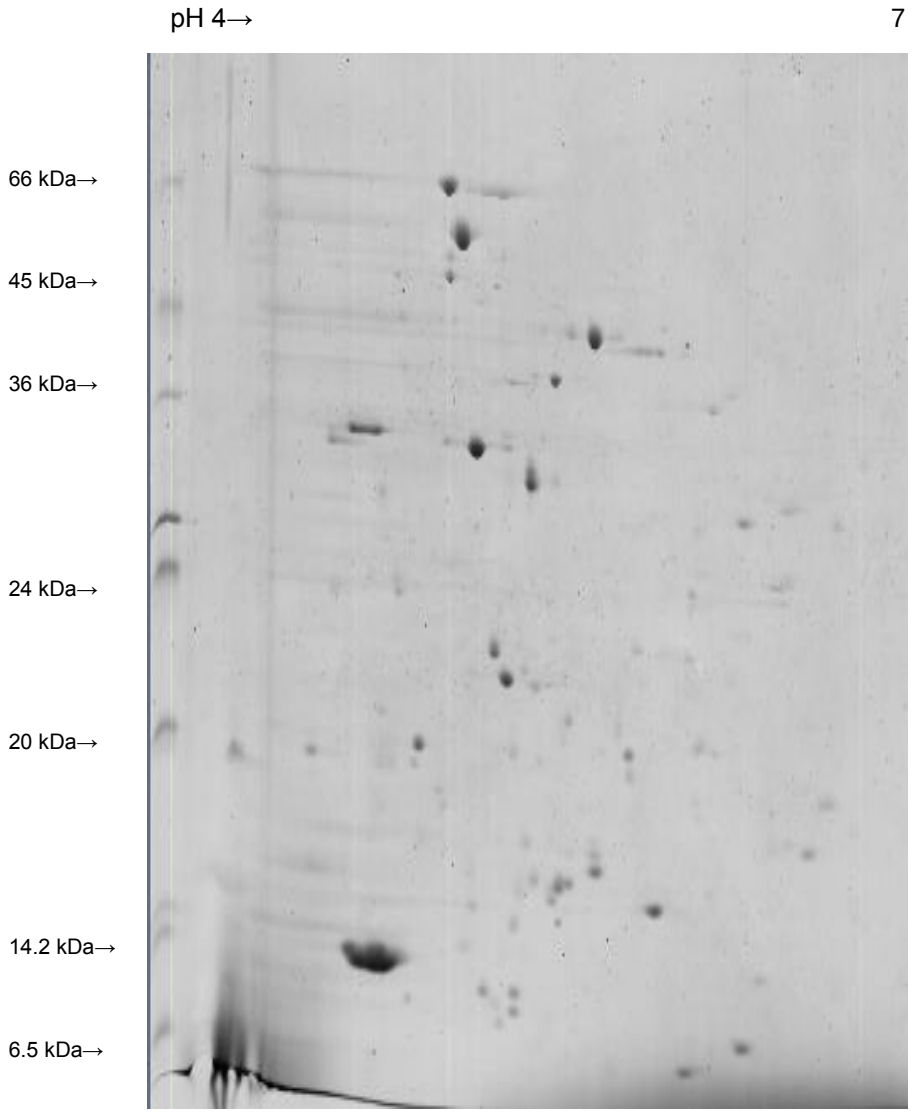


Figure 21.3: Proteome map no.3 of *E. cancerogenus* grown in mucin Type II enriched media

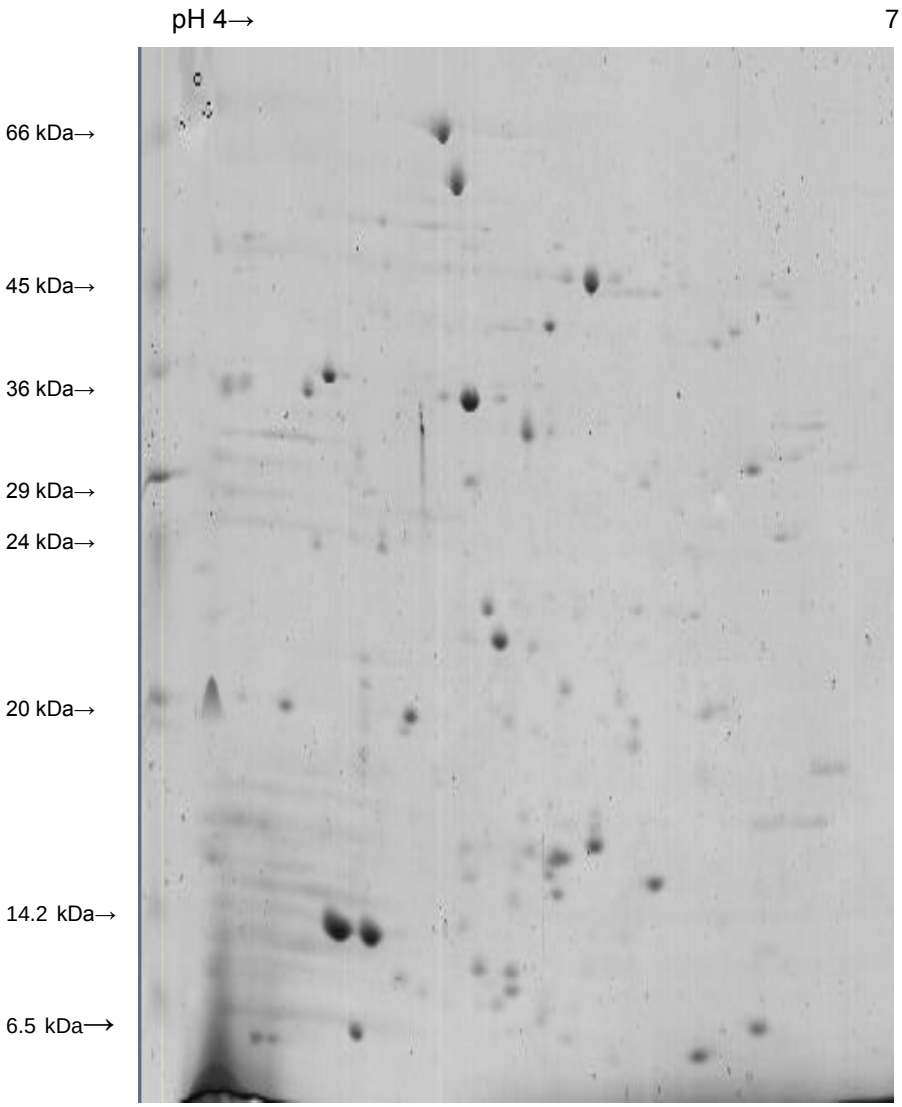


Figure 21.4: Proteome map no.4 of *E. cancerogenus* grown in mucin Type II enriched media

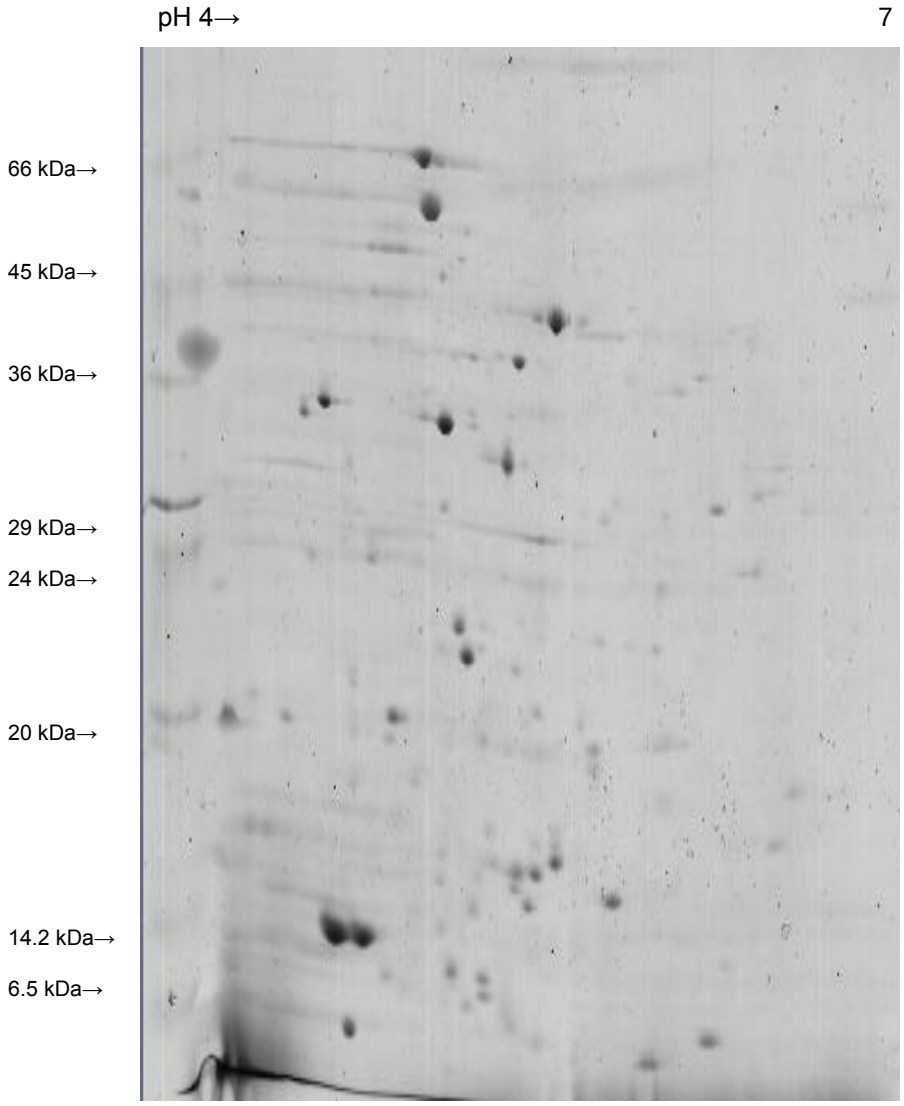


Figure 21.5: Proteome map no.5 of *E. cancerogenus* grown in mucin Type II enriched media

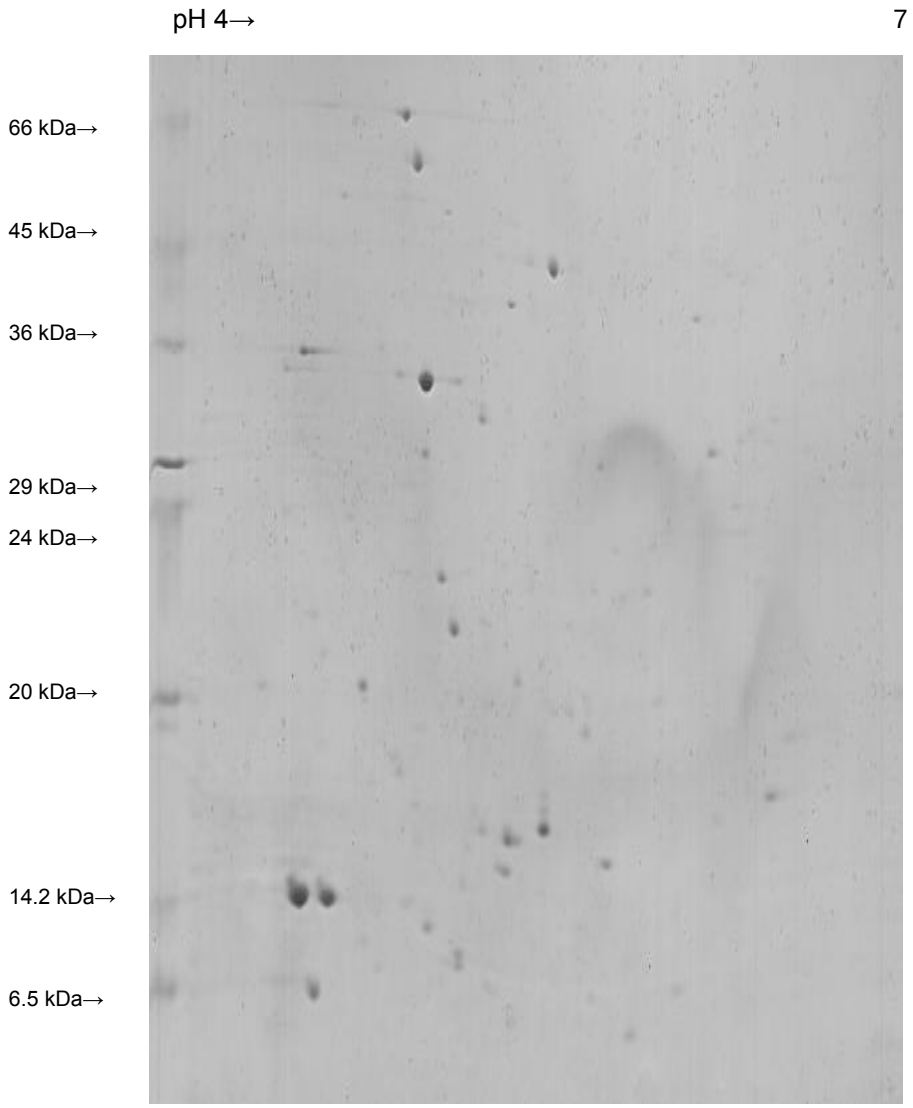


Figure 21.6: Proteome map no.6 of *E. cancerogenus* grown in mucin Type II enriched media

Figure 22: Proteome maps of *E. cancerogenus* grown in semi-defined media enriched with mucin Type III (using pH 4-7 IEF strips)

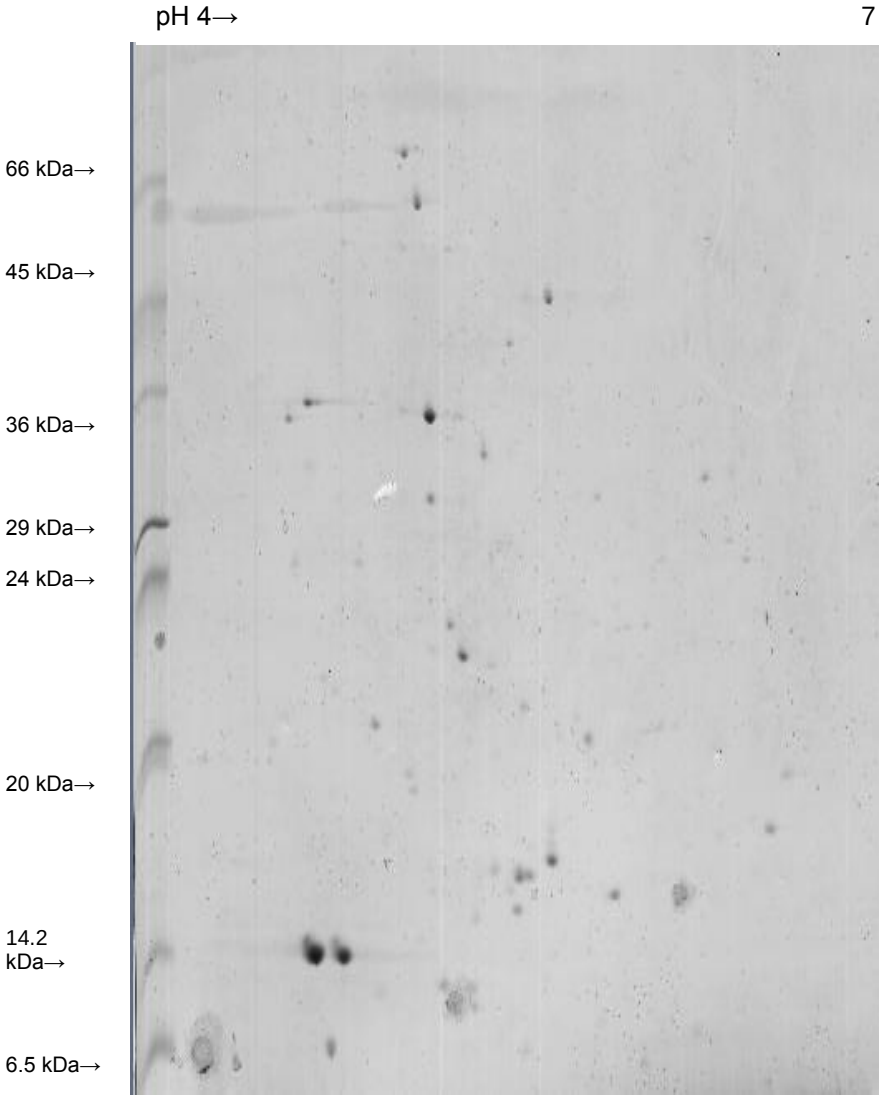


Figure 22.1: Proteome map no.1 of *E. cancerogenus* grown in mucin Type III enriched media

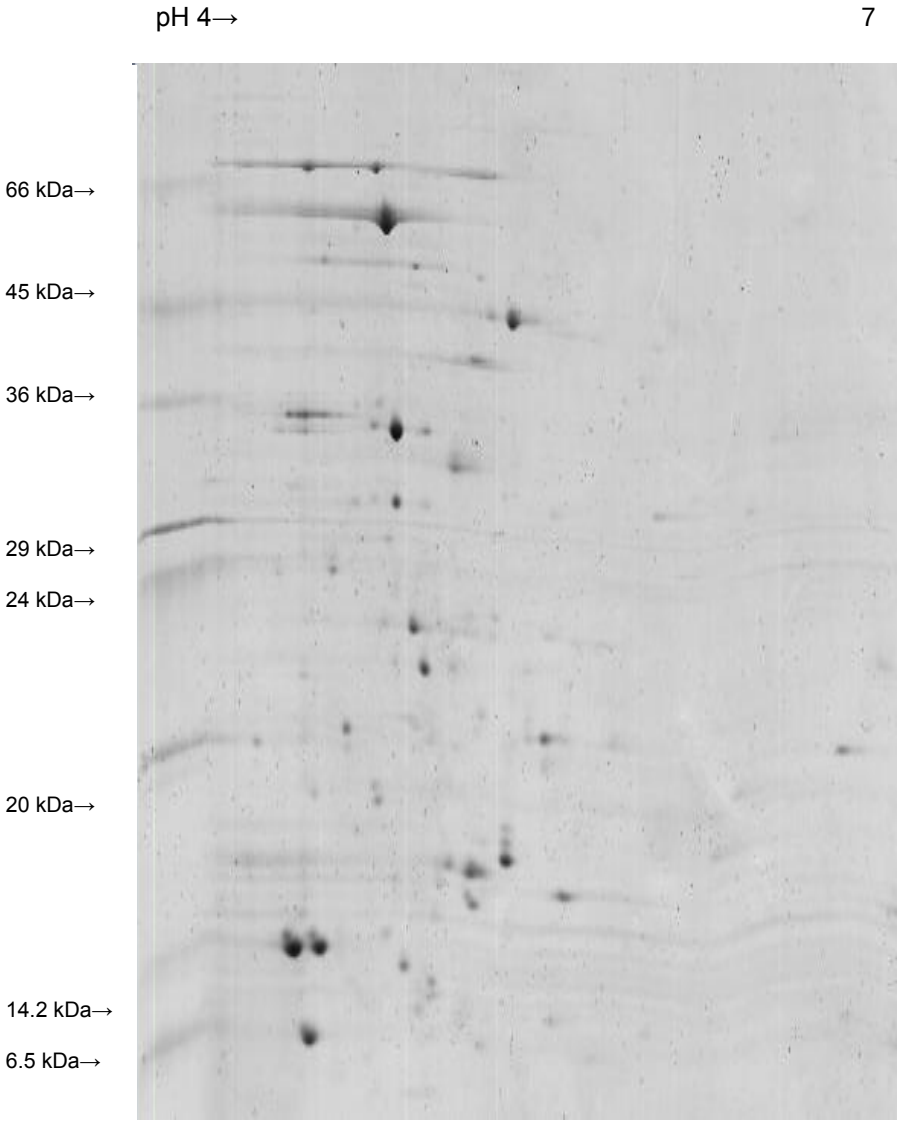


Figure 22.2: Proteome map no.2 of *E. cancerogenus* grown in mucin Type III enriched media

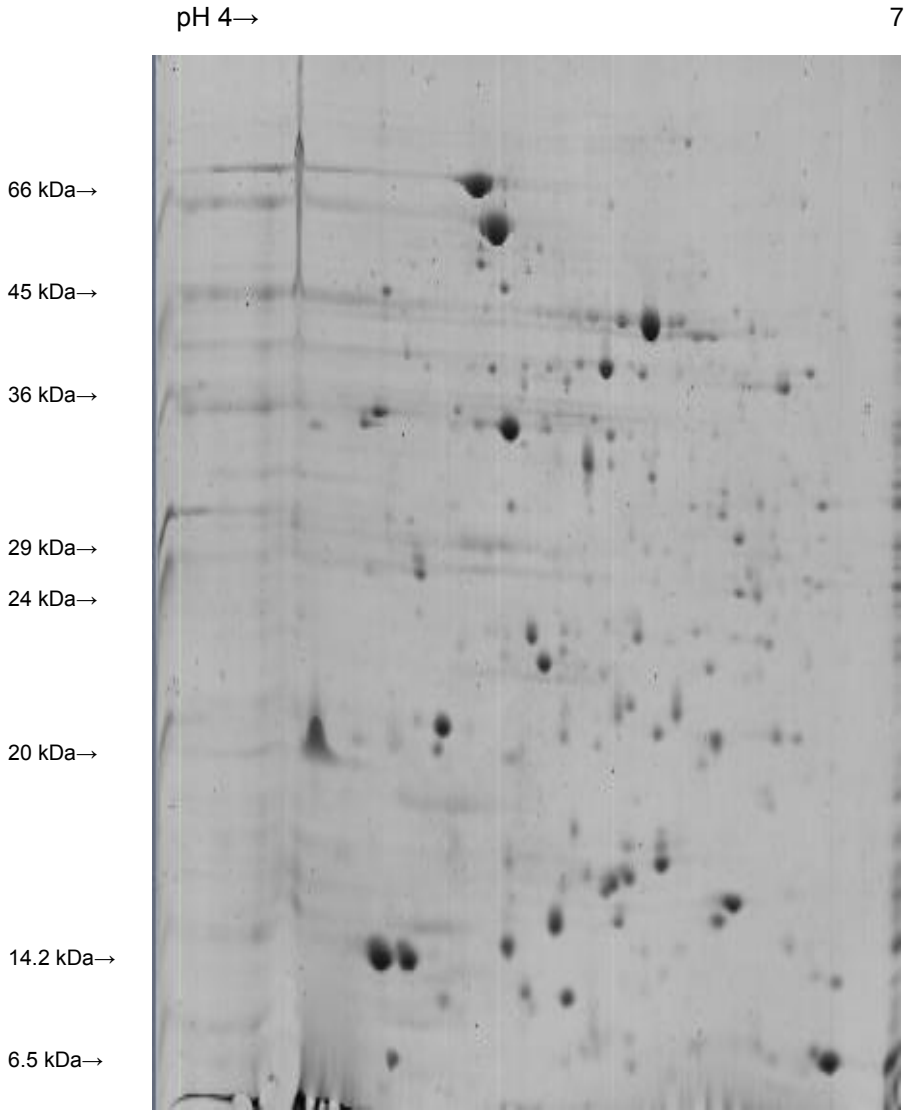


Figure 22.3: Proteome map no.3 of *E. cancerogenus* grown in mucin Type III enriched media

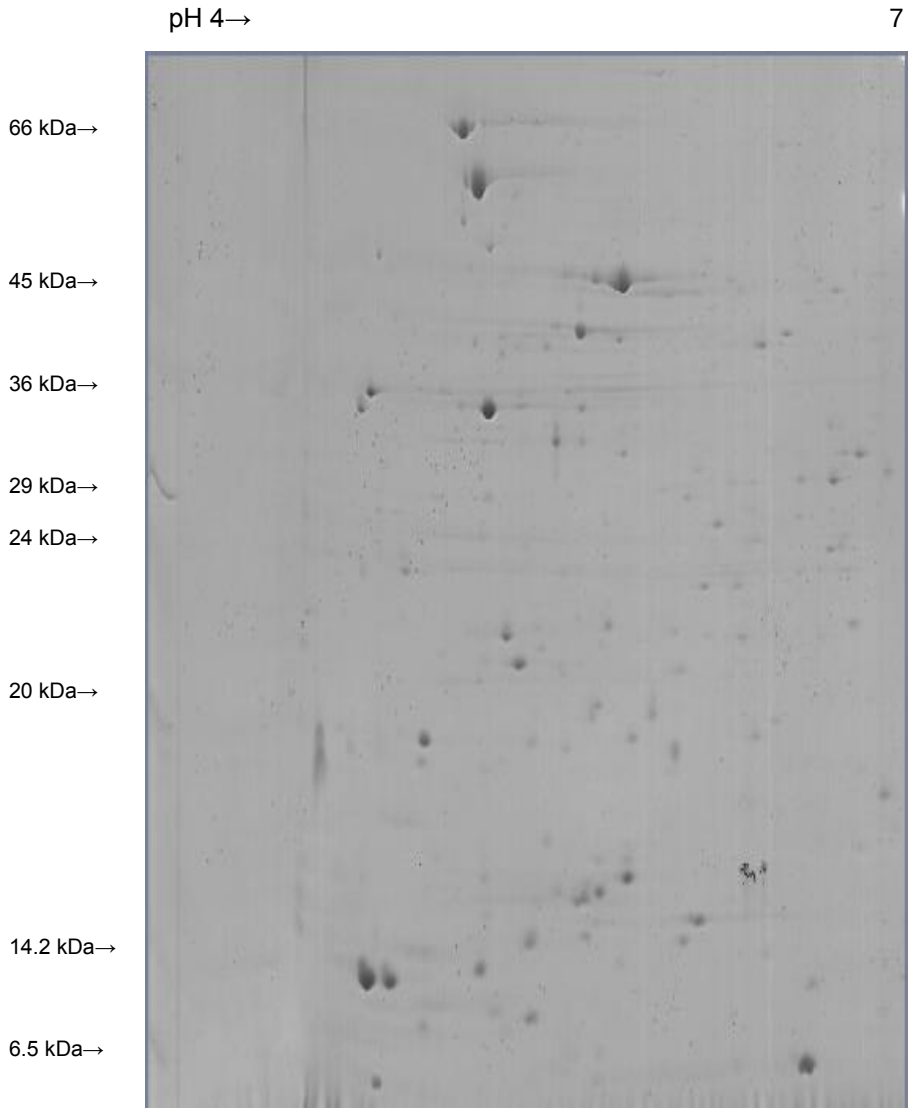


Figure 22.4: Proteome map no.4 of *E. cancerogenus* grown in mucin Type III enriched media

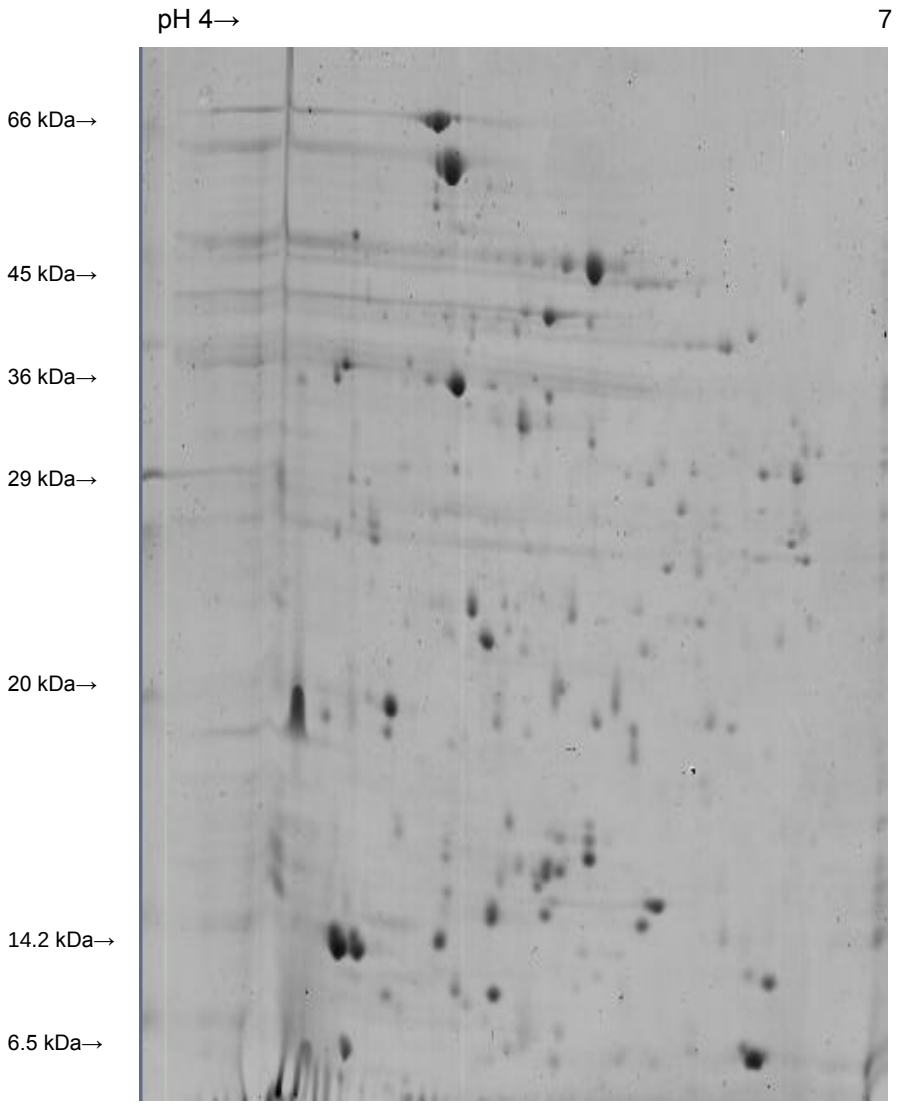


Figure 22.5: Proteome map no.5 of *E. cancerogenus* grown in mucin Type III enriched media

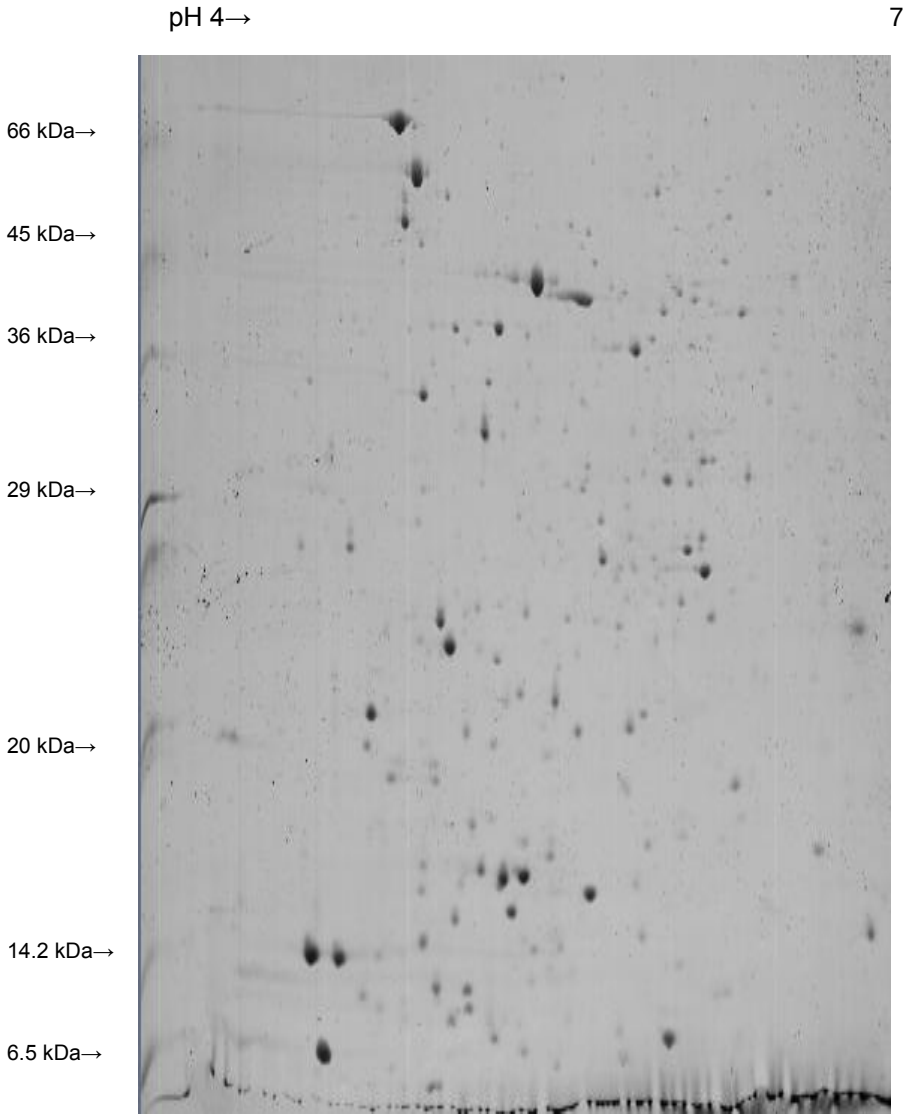


Figure 22.6: Proteome map no.6 of *E. cancerogenus* grown in mucin Type III enriched media

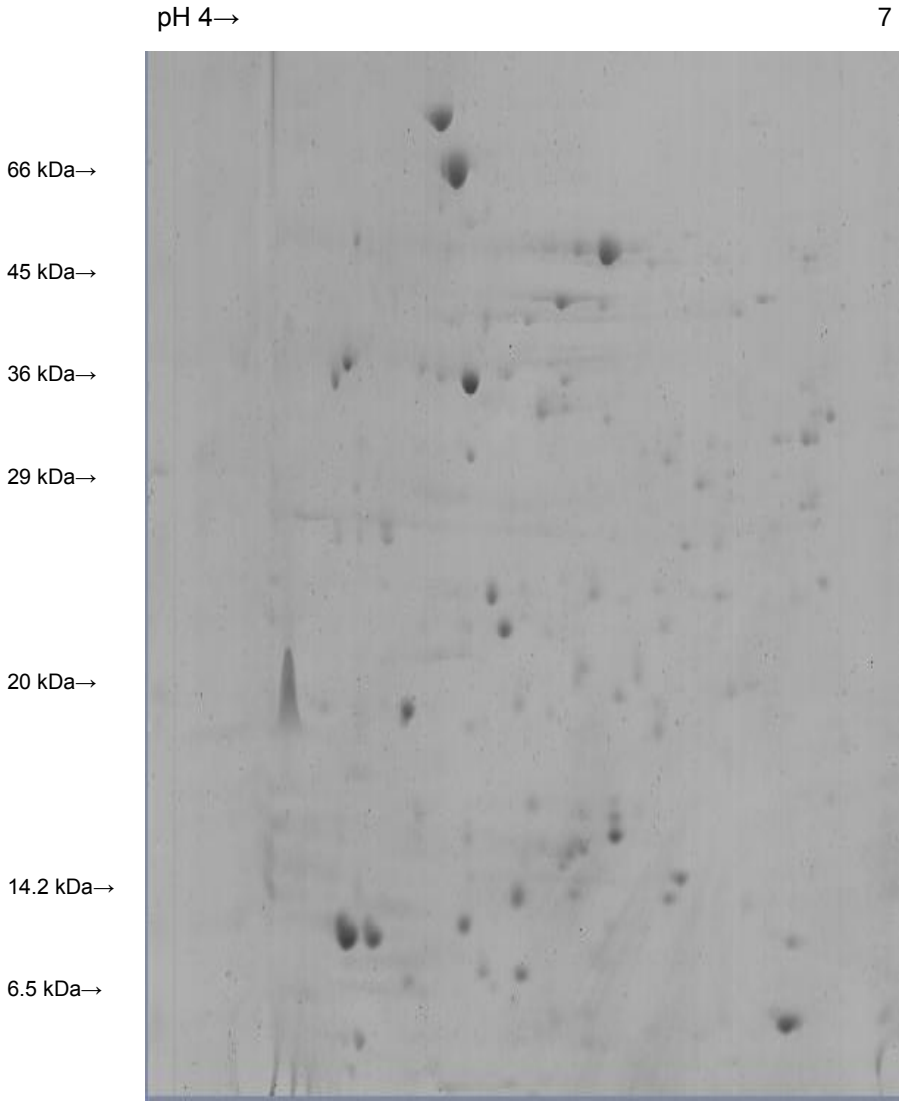


Figure 22.7: Proteome map no.7 of *E. cancerogenus* grown in mucin Type III enriched media

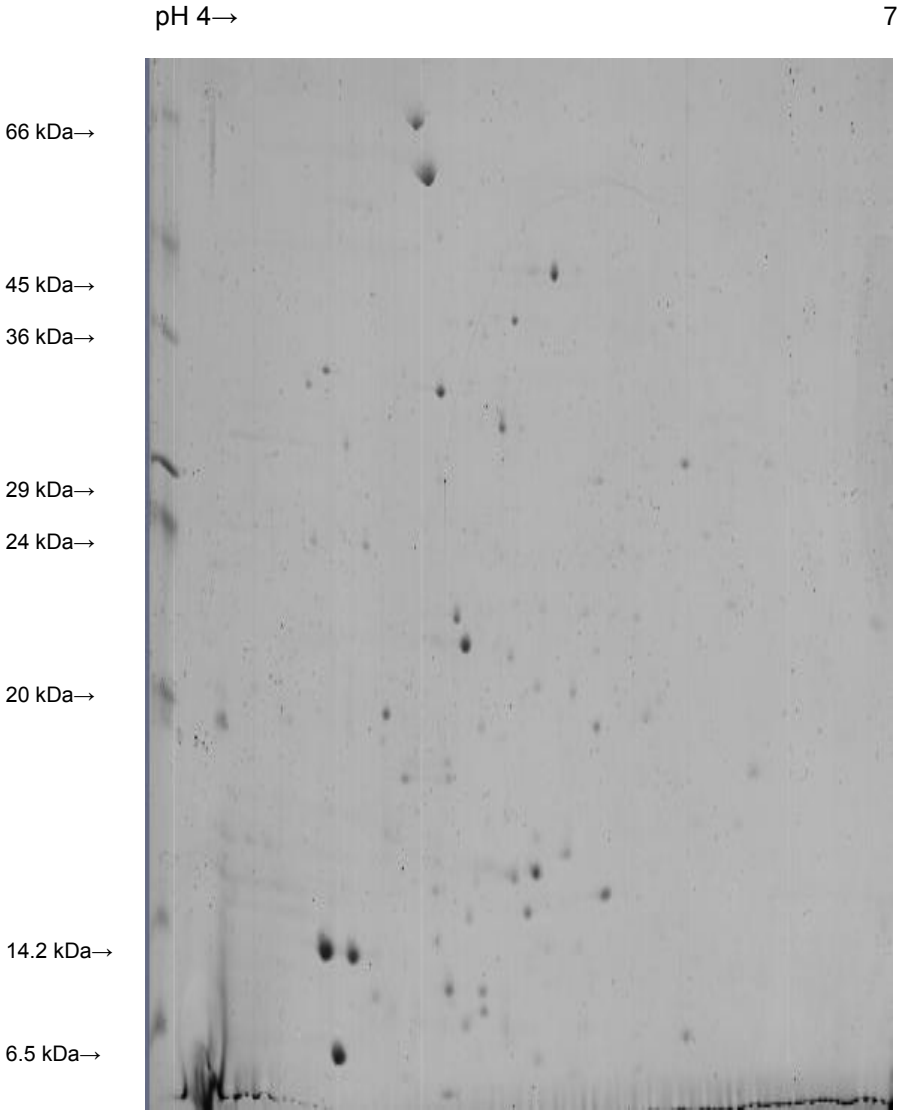


Figure 22.8: Proteome map no.8 of *E. cancerogenus* grown in mucin Type III enriched media

10.6.2 Sets of replicate 2DE gels of proteins from *B. fragilis*

Multiple sets of gels of proteins from *B. fragilis* grown in media without mucin, media with mucin Type II and media with mucin Type III (IEF strip pH range 4-7, processing volume: 100 mL, harvesting OD at 600 nm: 0.7) (Figs. 23-25).

Figure 23: Proteome maps of *B. fragilis* grown in semi-defined media without mucin (control media; using pH 4-7 IEF strips)

.

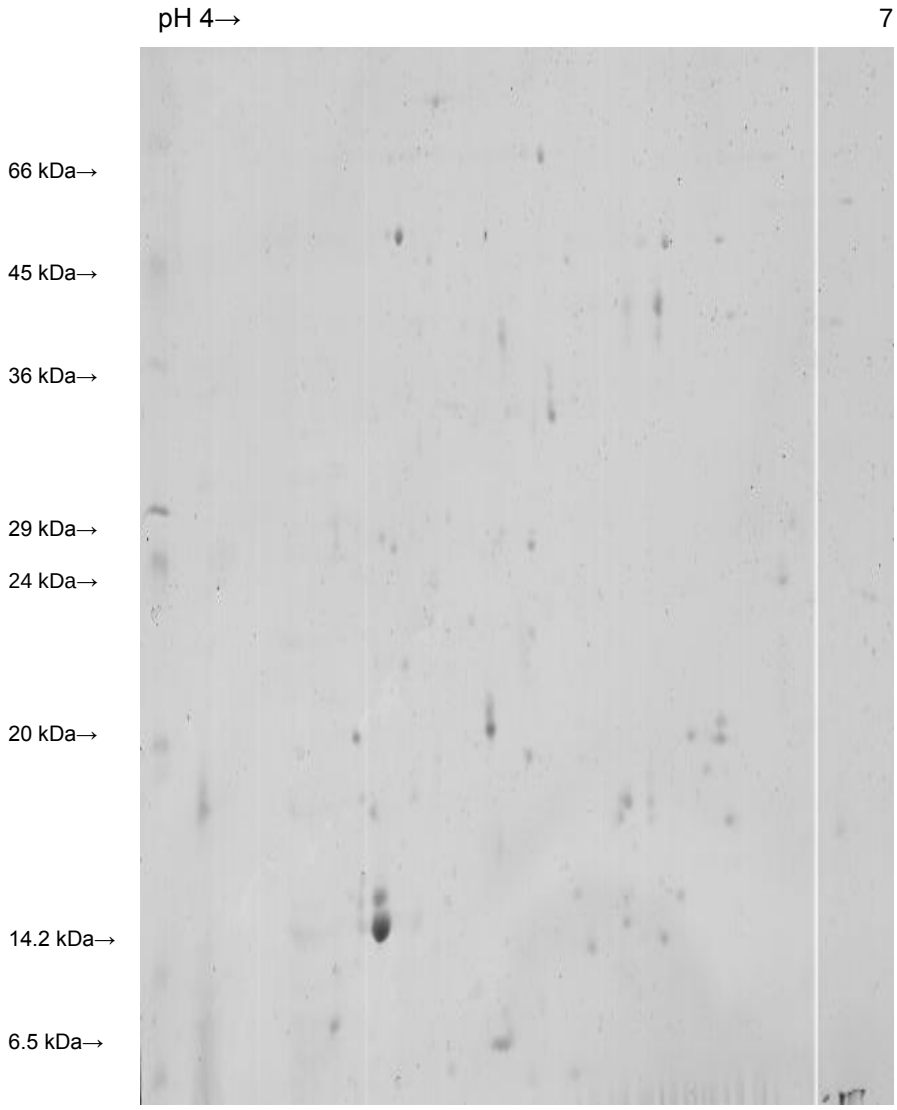


Figure 23.1 Proteome map no.1 of *B. fragilis* grown in semi-defined media without mucin

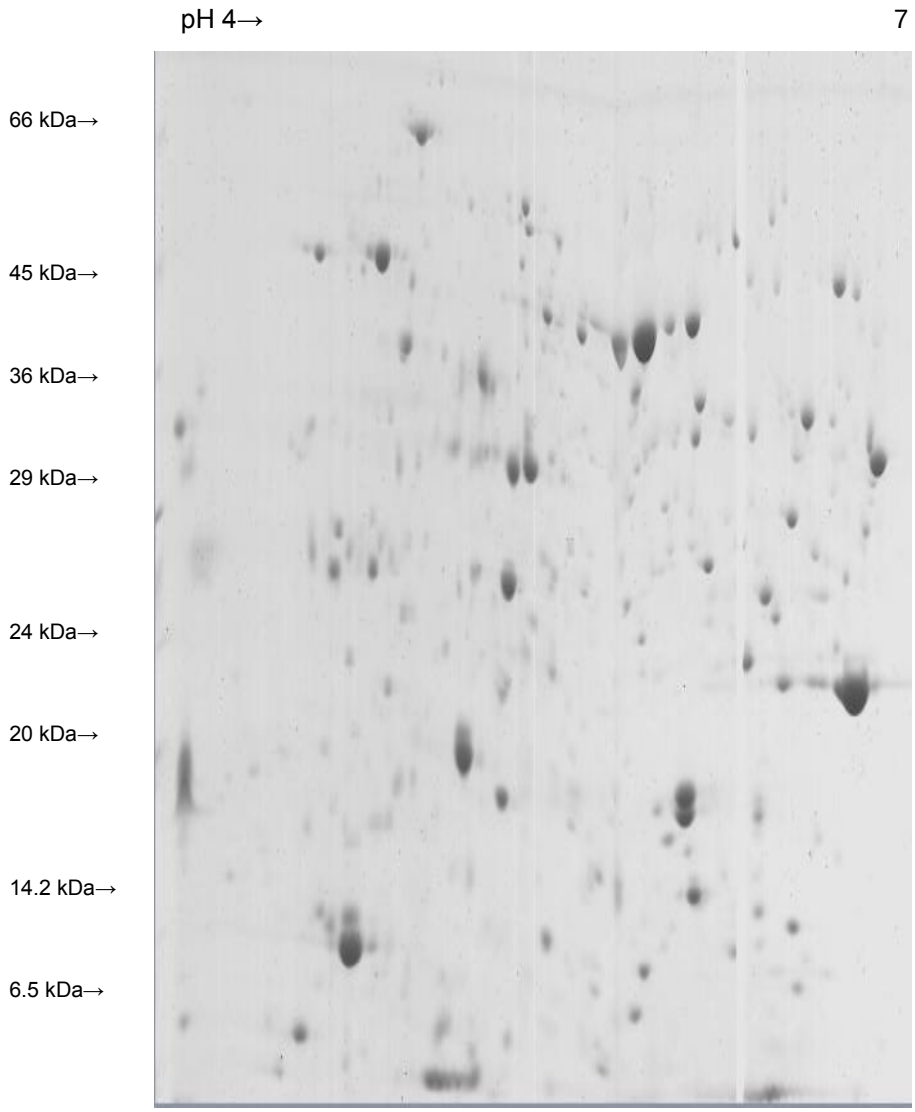


Figure 23.2: Proteome map no.2 of *B. fragilis* grown in semi-defined media without mucin

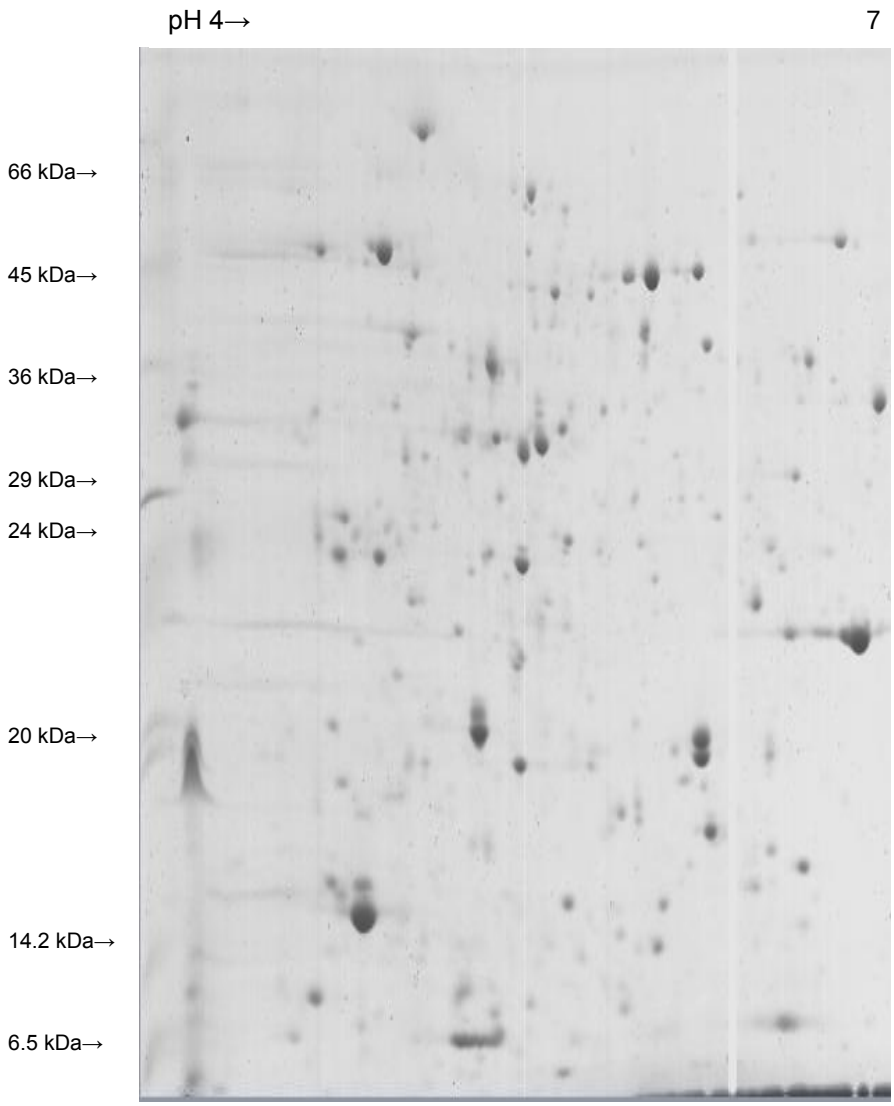


Figure 23.3: Proteome map no.3 of *B. fragilis* grown in semi-defined media without mucin

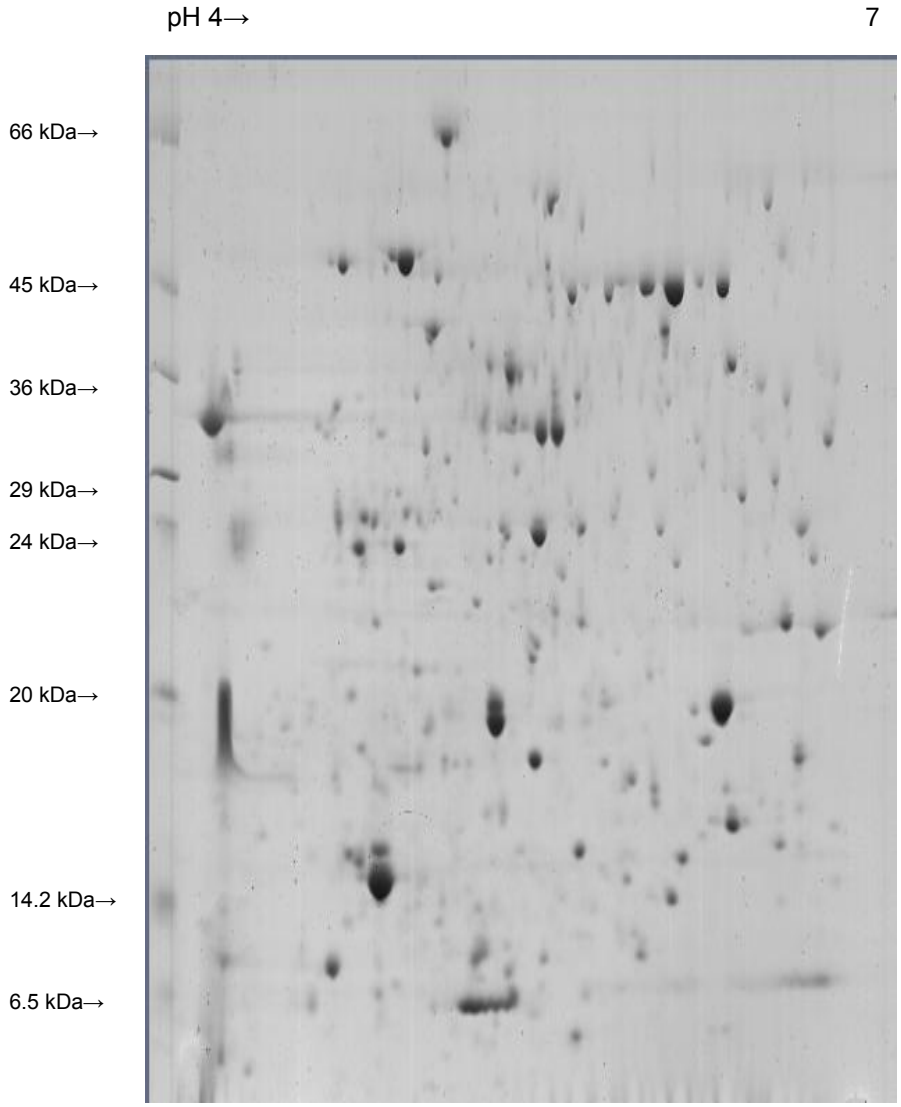


Figure 23.4: Proteome map no.4 of *B. fragilis* grown in semi-defined media without mucin

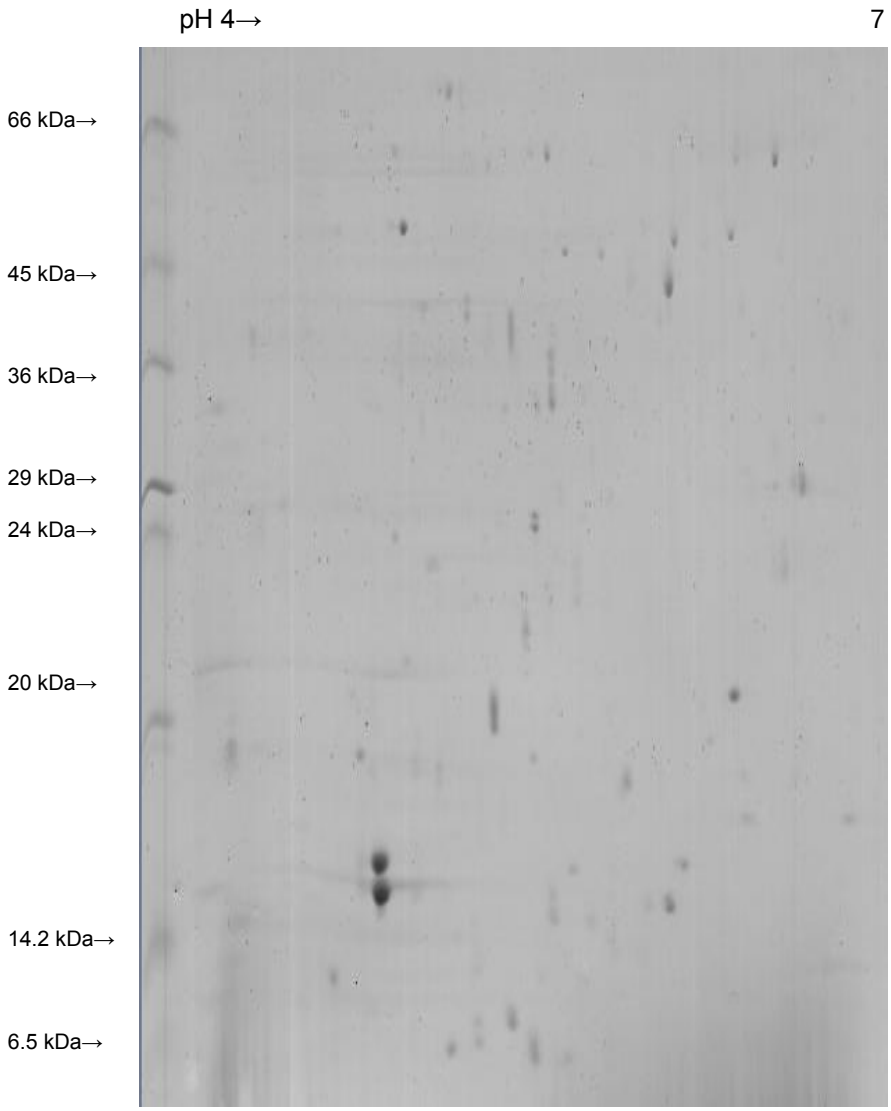


Figure 23.5: Proteome map no.5 of *B. fragilis* grown on semi-defined media without mucin

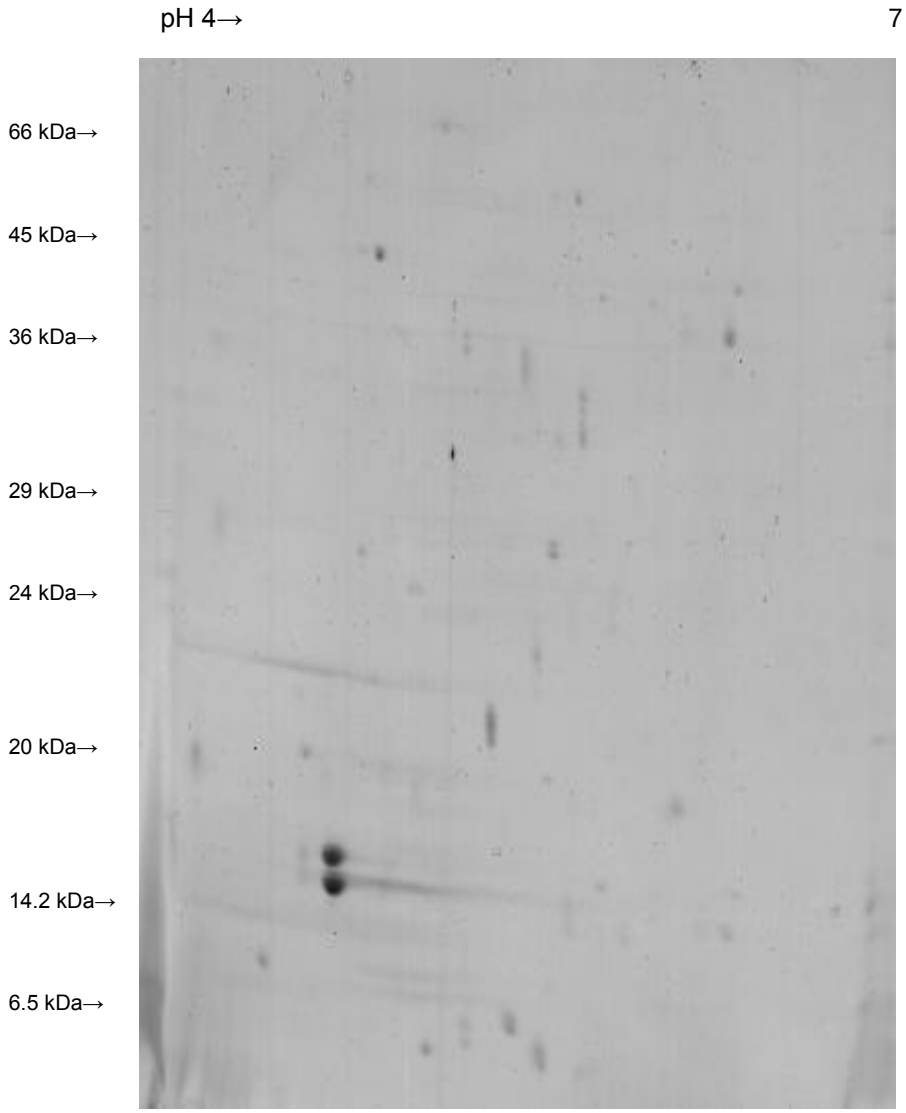


Figure 23.6: Proteome map no.6 of *B. fragilis* grown in semi-defined media without mucin

Figure 24: Proteome maps of *B. fragilis* grown in media containing mucin Type II (using pH 4-7 IEF strips).

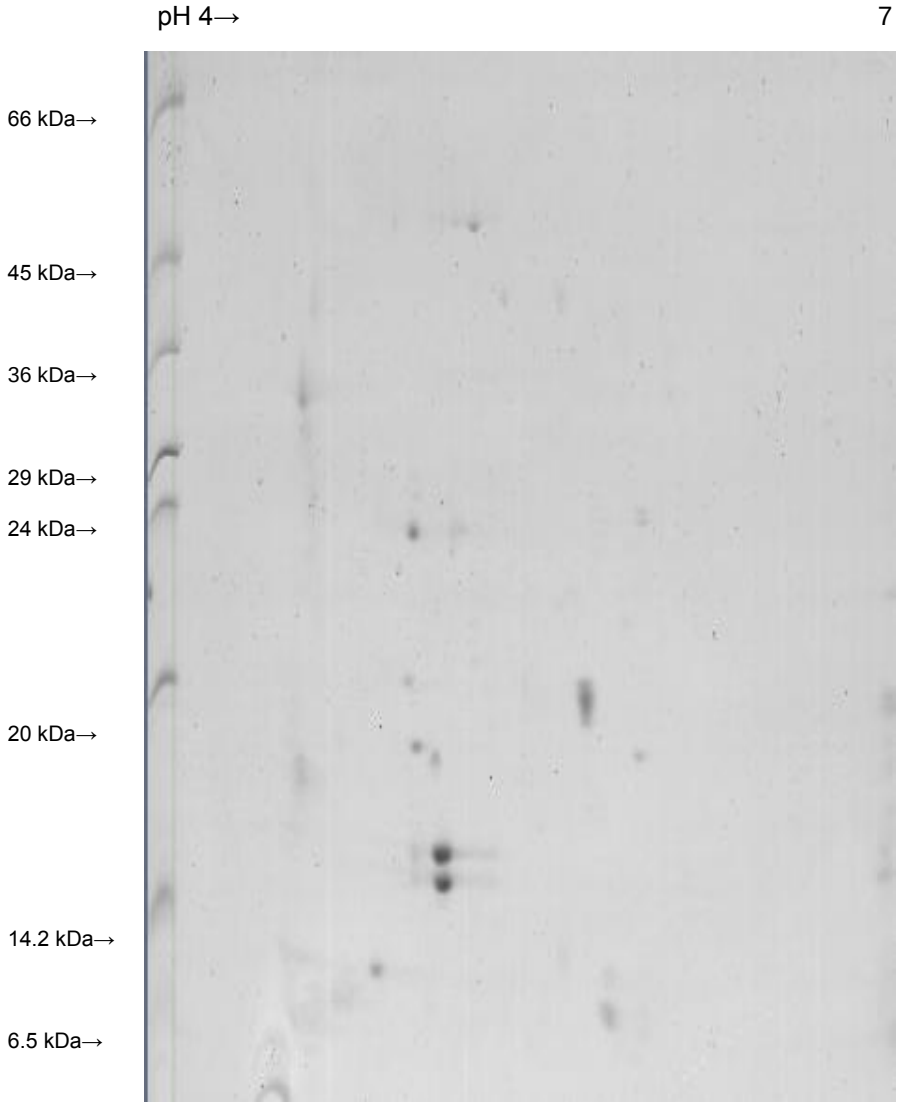


Figure 24.1: Proteome map no.1 of *B. fragilis* grown in mucin Type II enriched media

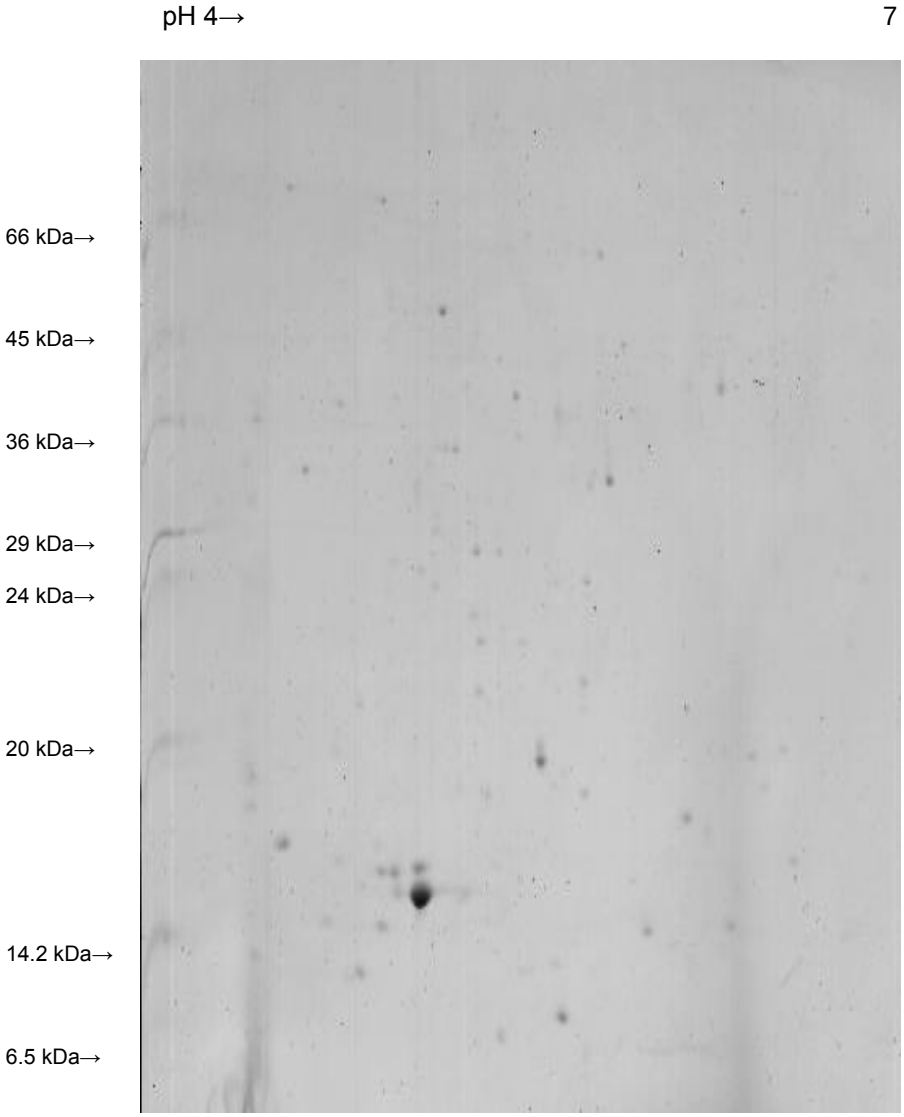


Figure 24.2: Proteome map no.2 of *B. fragilis* grown in mucin Type II enriched media

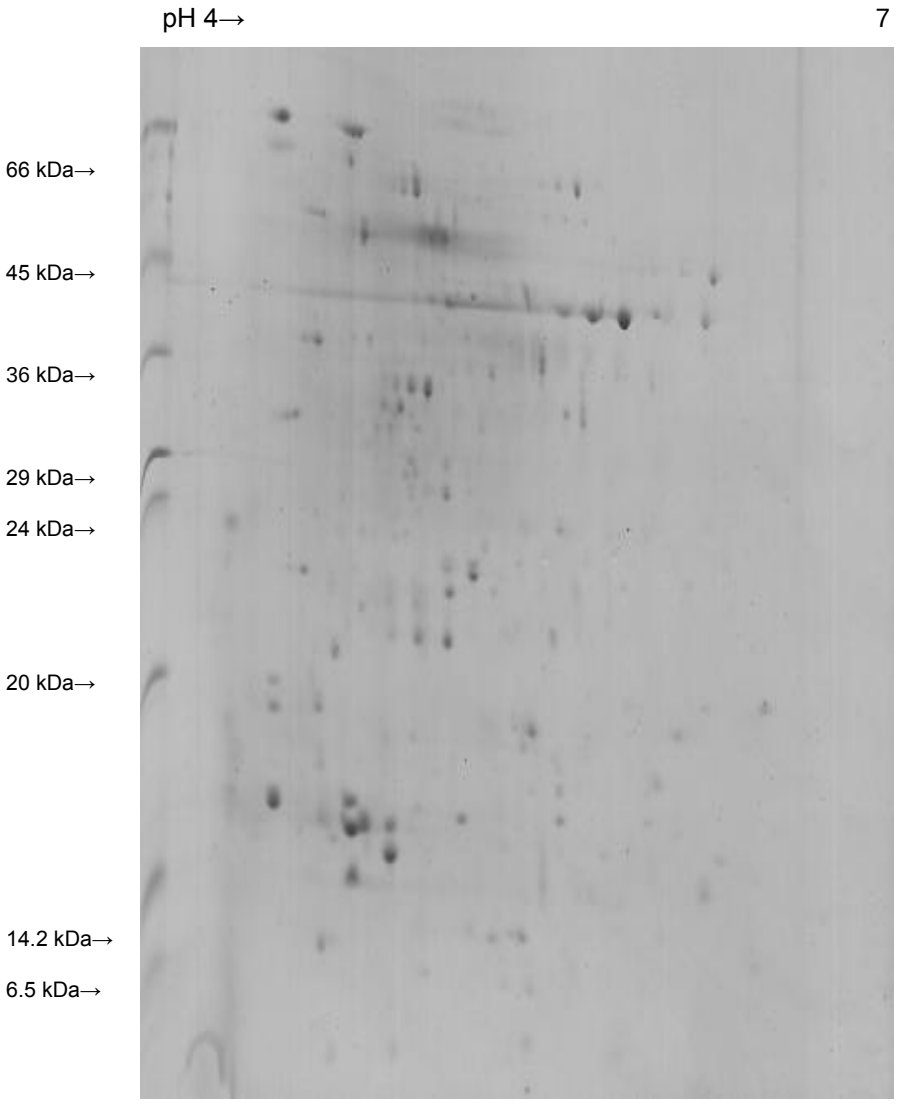


Figure 24.3: Proteome map no.3 of *B. fragilis* grown in mucin Type II enriched media

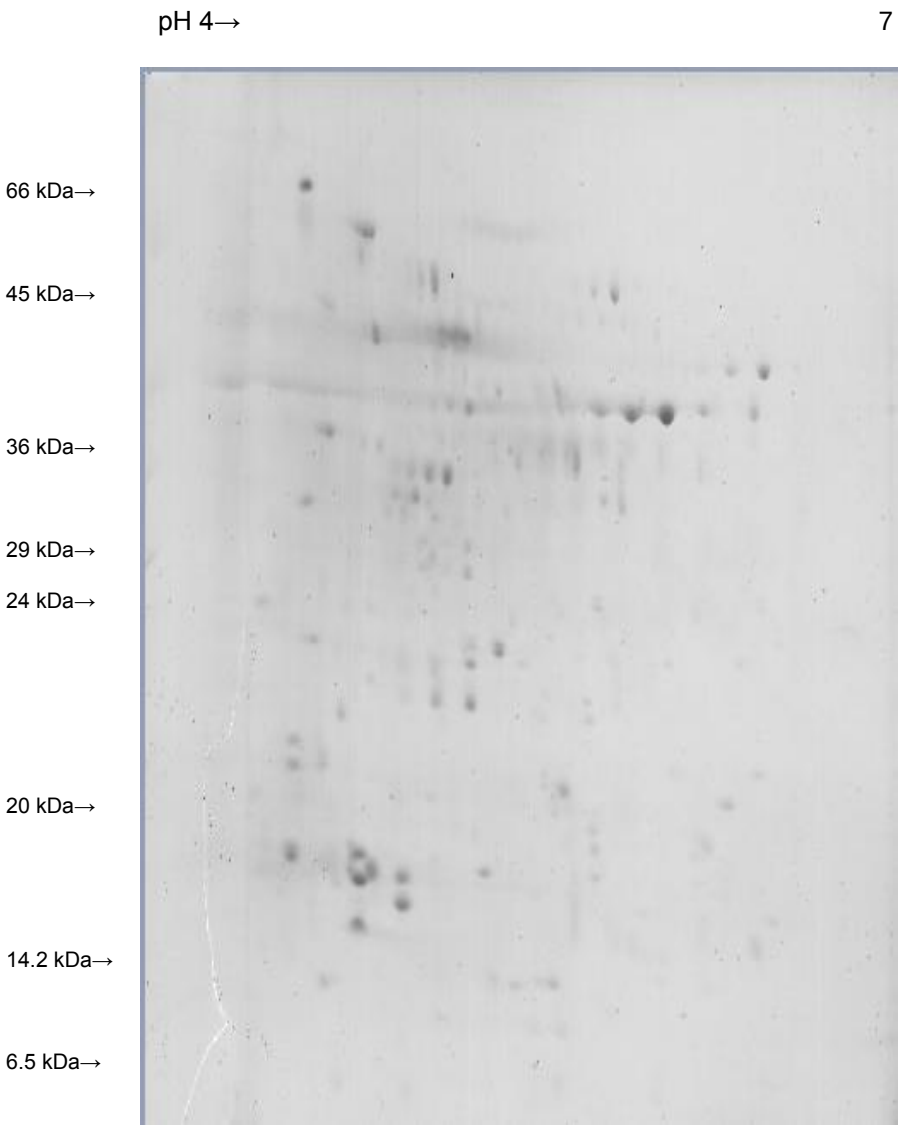


Figure 24.4: Proteome map no.4 of *B. fragilis* grown in mucin Type II enriched media

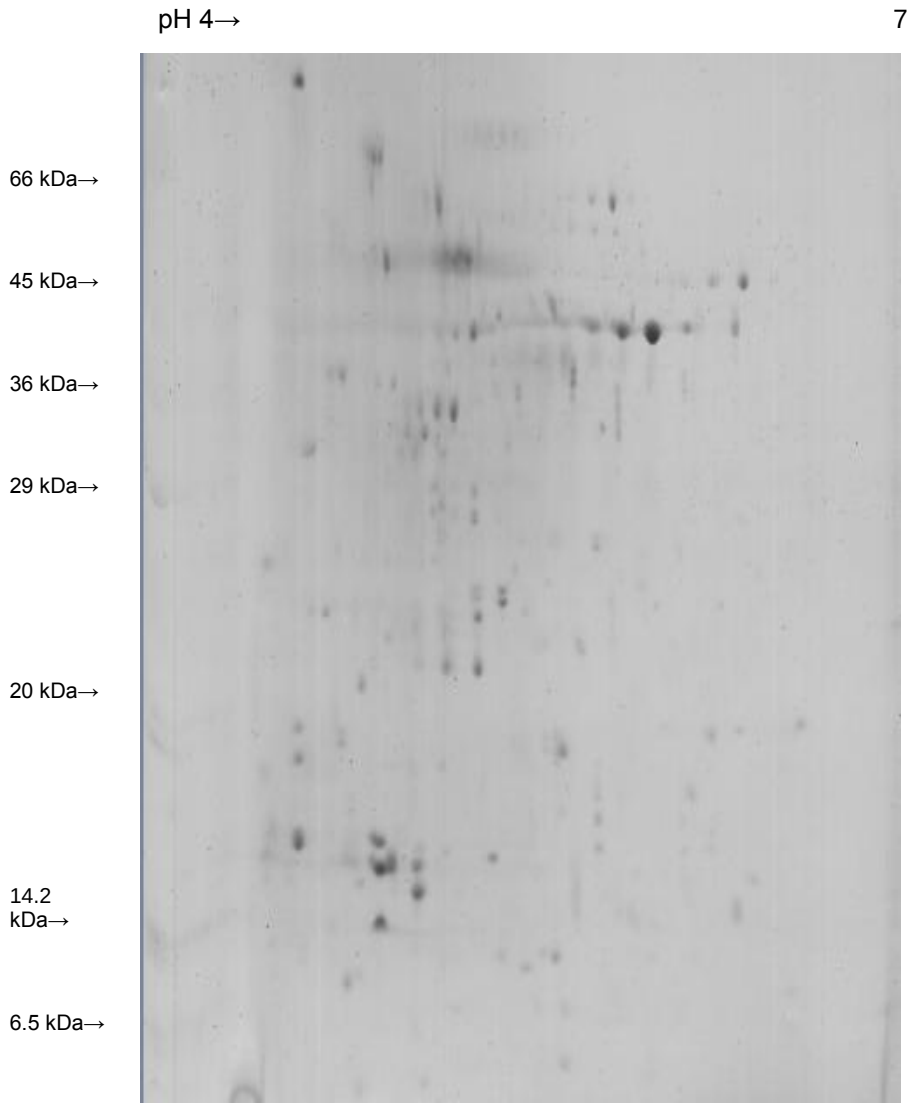


Figure 24.5: Proteome map no.5 of *B. fragilis* grown in mucin Type II enriched media

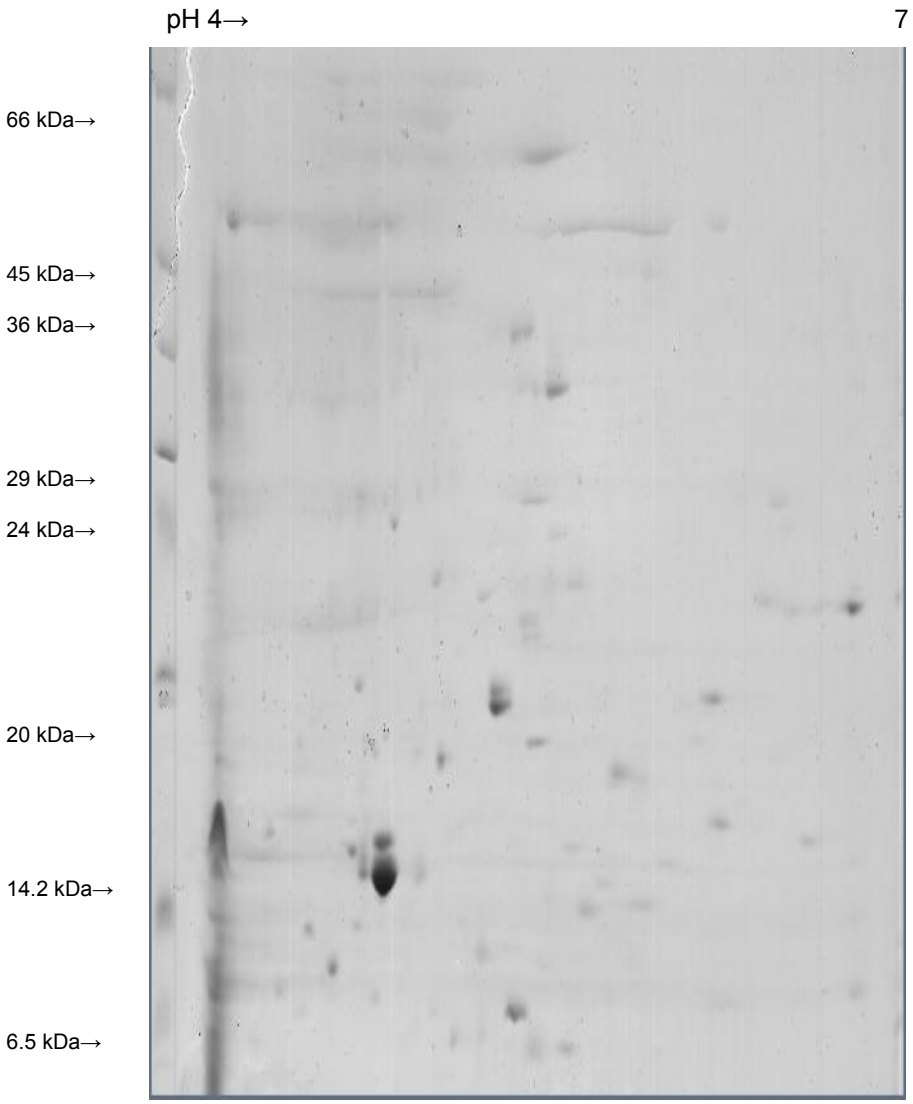


Figure 24.6: Proteome map no.6 of *B. fragilis* grown in mucin Type II enriched media

Figure 25: Proteome maps of *B. fragilis* grown in semi-defined media enriched with mucin Type III (pH 4-7 IEF strips)

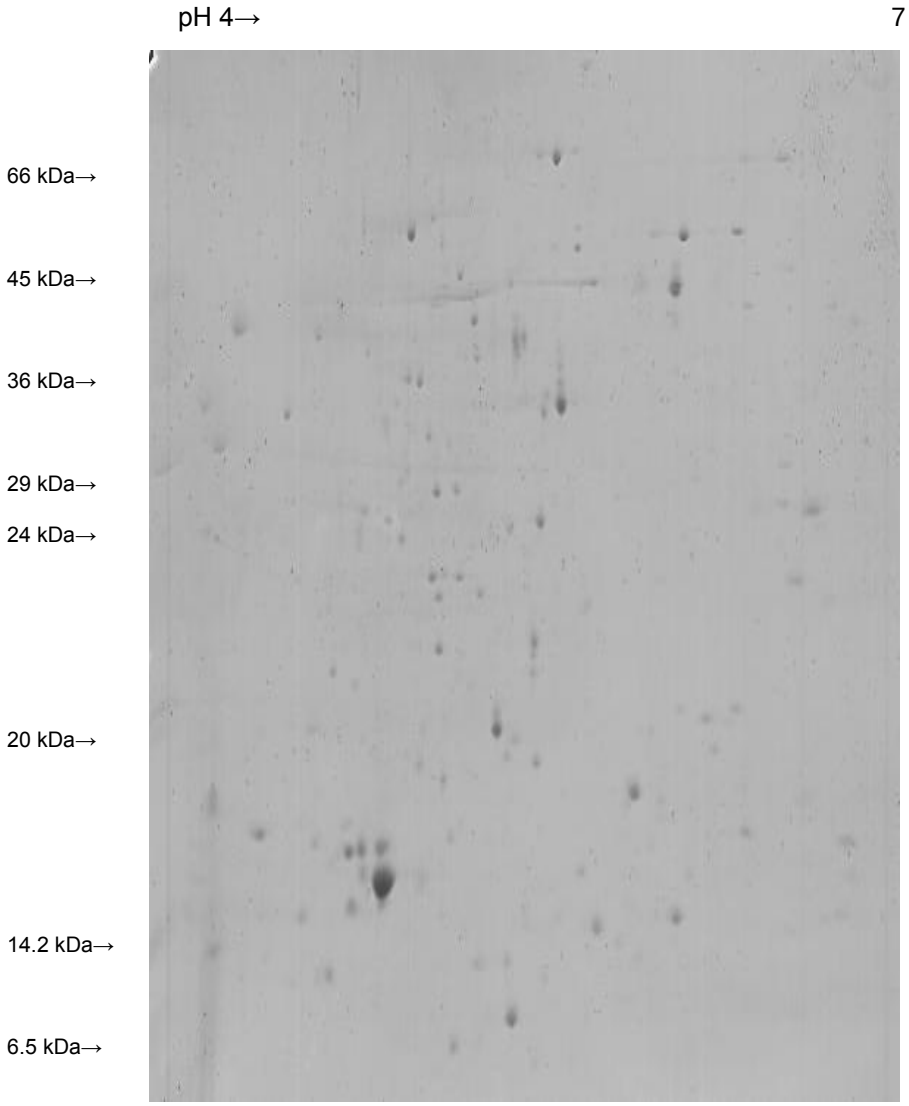


Figure 25.1: Proteome map no.1 of *B. fragilis* grown in mucin Type III enriched media

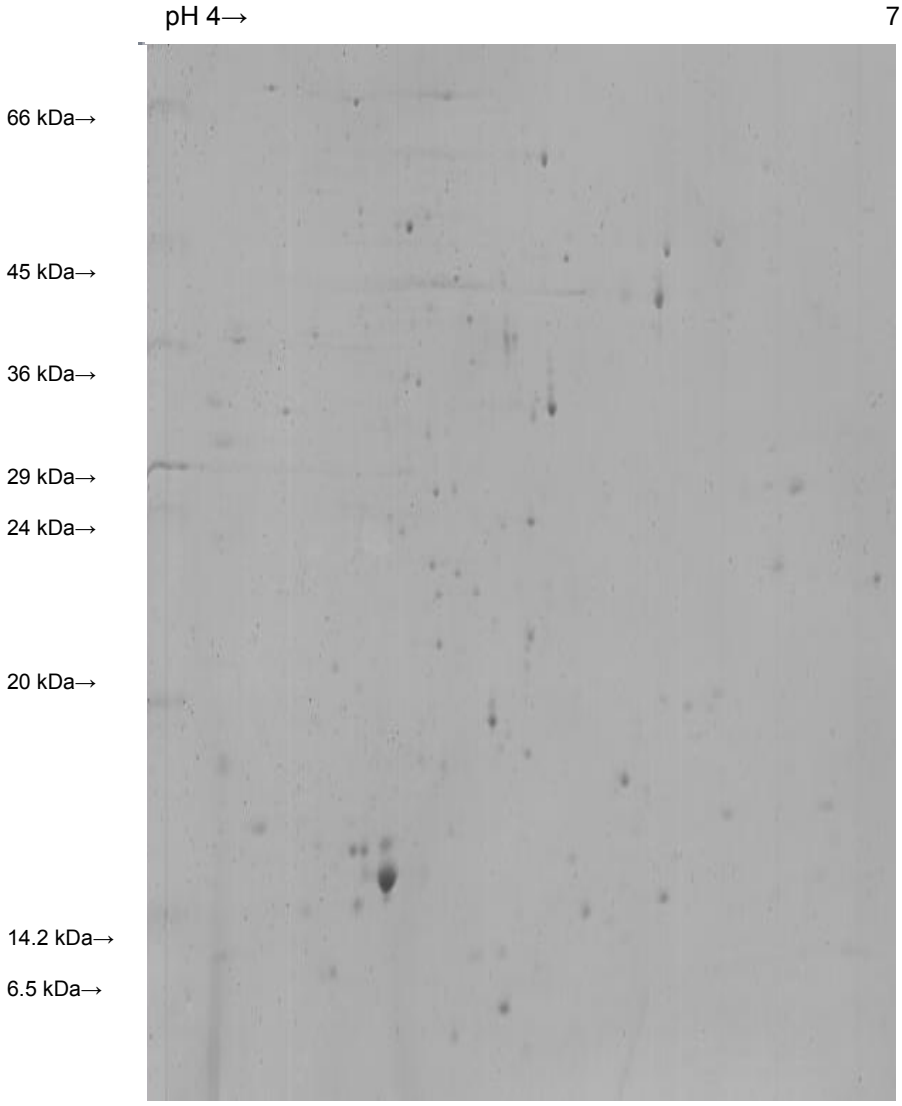


Figure 25.2: Proteome map no.2 of *B. fragilis* grown in mucin Type III enriched media

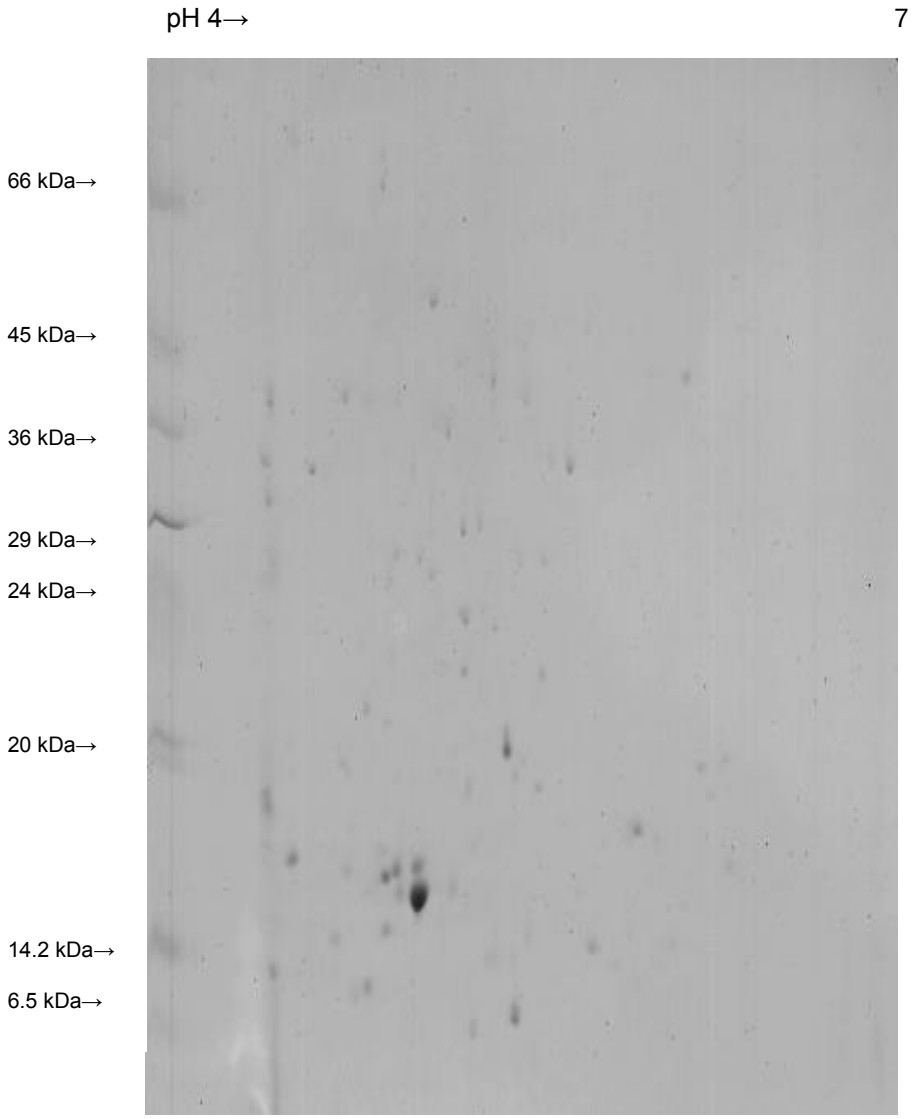


Figure 25.3: Proteome map no.3 of *B. fragilis* grown in mucin Type III enriched media

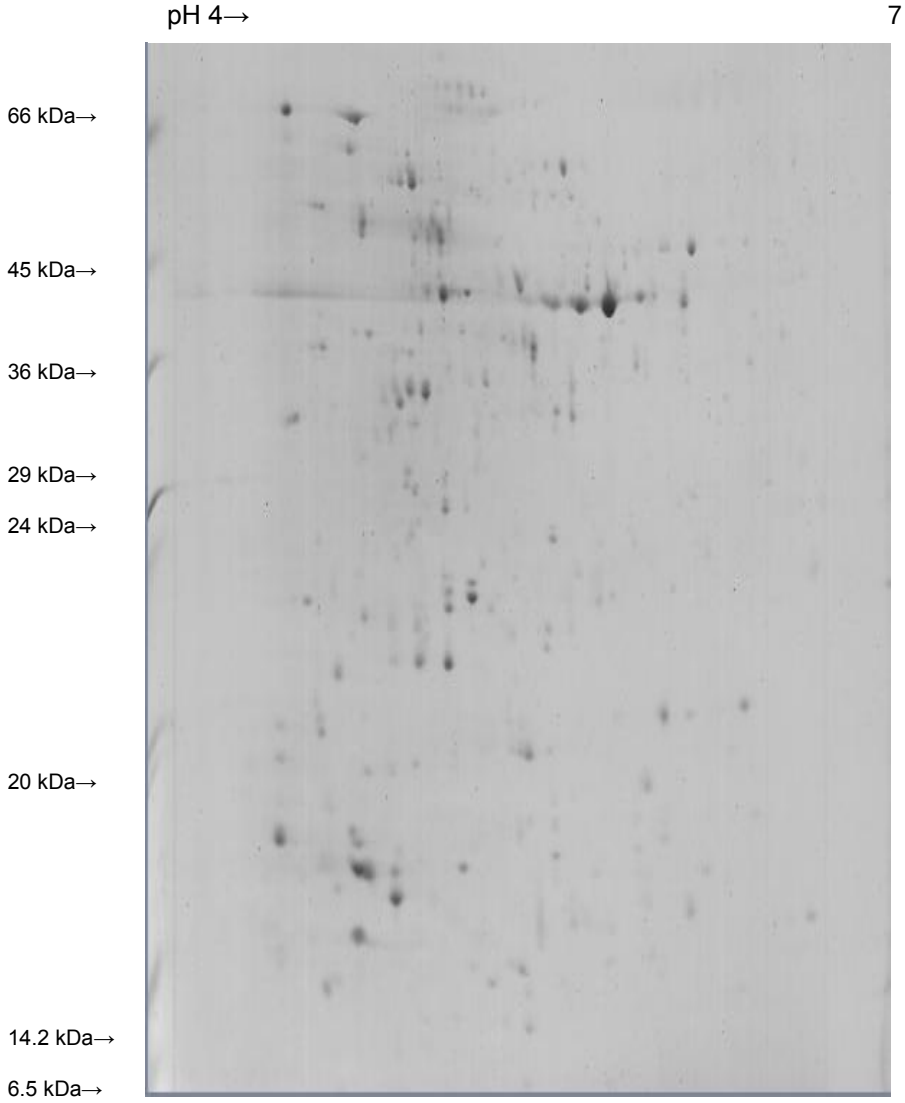


Figure 25.4: Proteome map no.4 of *B. fragilis* grown in mucin Type III enriched media

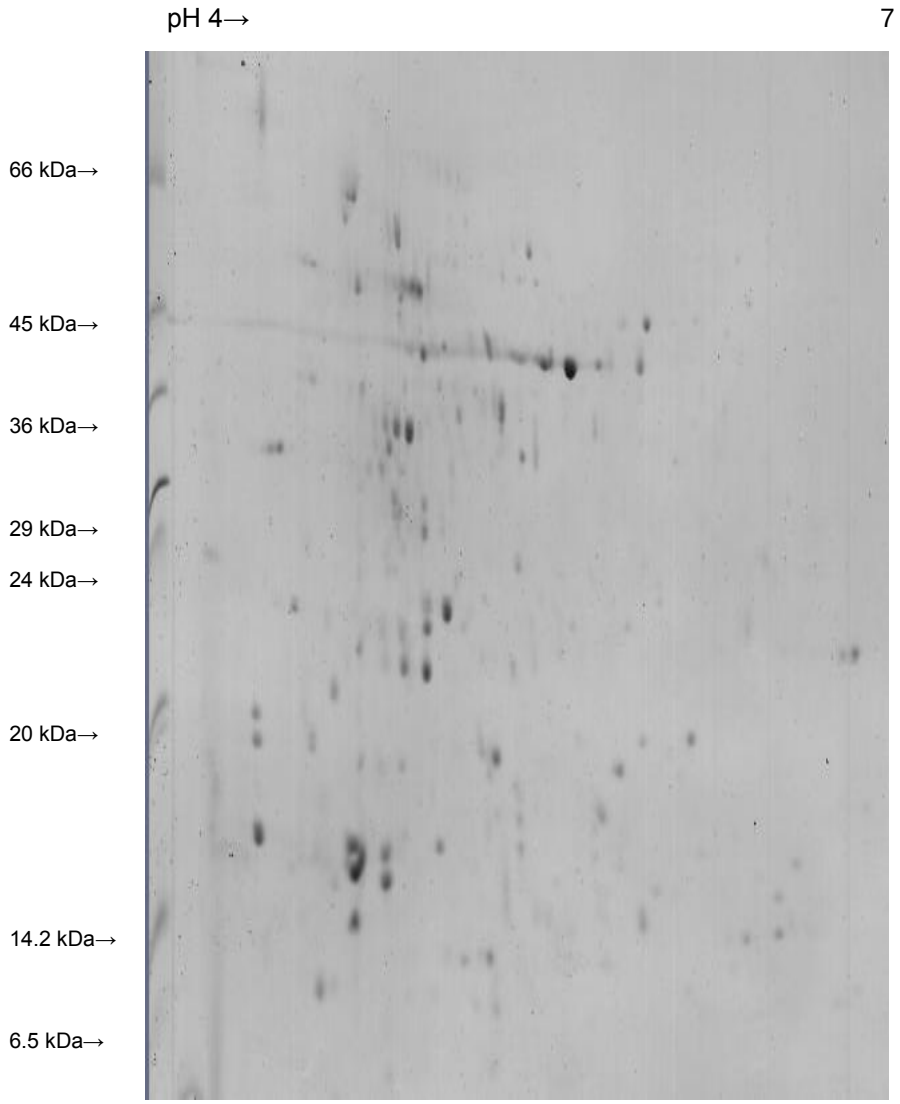


Figure 25.5: Proteome map no.5 of *B. fragilis* grown in mucin Type III enriched media

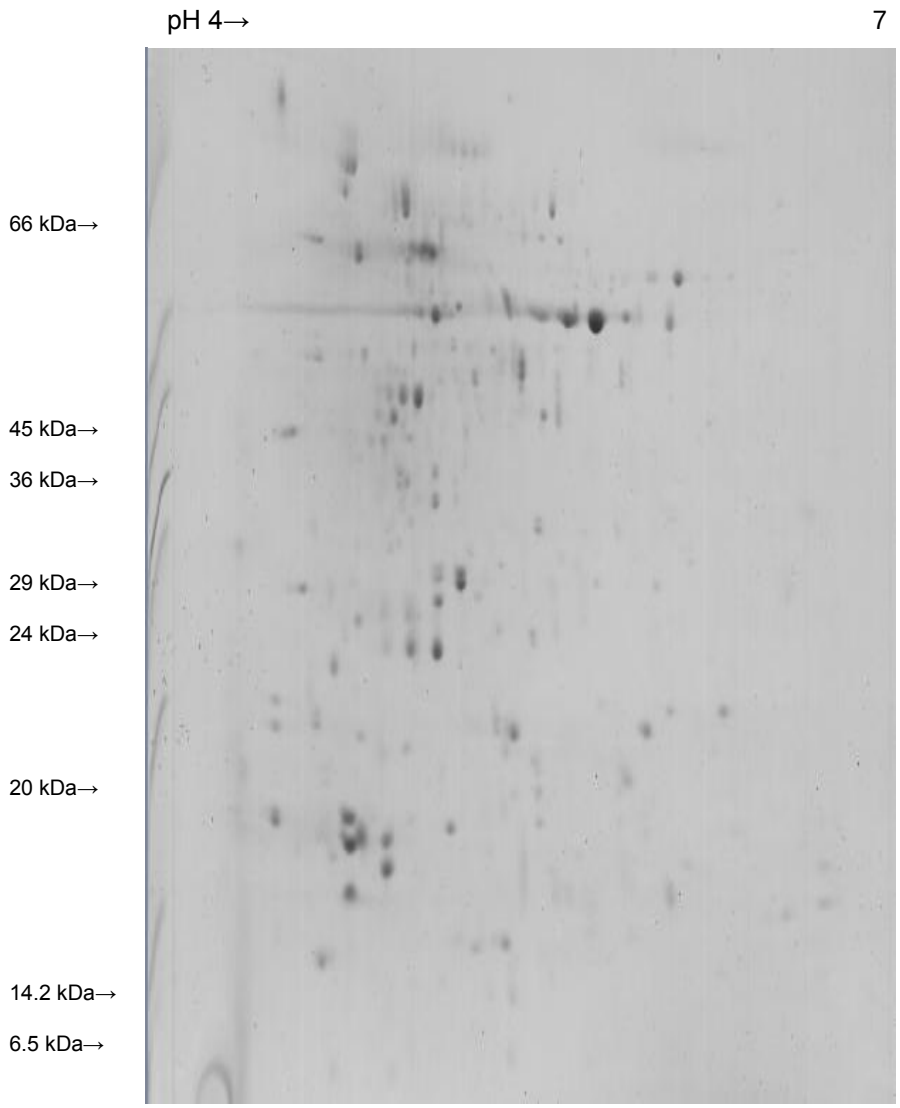


Figure 25.6: Proteome map no.6 of *B. fragilis* grown in mucin Type III enriched media

10.7 Analysis of 2DE gels

10.7.1 Analysis of 2DE gels using Bio-Rad PDQuest software

PDQuest software was used to analyse the spots present in the three different conditions. A reference master image was created which contained all the spots from all the three different conditions. Further landmarking and matching of gels was performed using the reference master gel as the template (Fig. 26.1, 26.3). The filename 'PROTBF 1' refers to the 2D gel analysis in *E. cancerogenus* and the filename 'PROTEOMICS 3' refers to the 2D gel analysis in *B. fragilis*. Refer to Appendix F for details.

The spot review tool was used to generate bar charts showing the expression of all the spots included in the reference master gel (Appendix F1). The spots are assigned SSP numbers and the graphs show the level of expression of a spot in all the individual gels or can be adjusted to show average values within replicate groups. Differentially expressed proteins were identified based on these expression patterns and cut out from the gel for further analysis. The analysis manager was used to perform statistical analysis on the spot data and the two main statistical tests performed were the Student t-test and the Mann-Whitney U test (Table 11).

Replicate groups used in analysis	Student t-test	Mann-Whitney U test	Organism
Non mucin v. mucin Type II enriched media	85/127	42/127	*E. cancerogenus*
Non mucin v. mucin Type III enriched media	32/65	33/65	*E. cancerogenus*
Non mucin v. mucin Type II enriched media	45/84	39/84	*B. fragilis*
Non mucin v. mucin Type III enriched media	66/94	28/94	*B. fragilis*

Table 11: Table represents the statistical test results for spot data in replicate groups using the Bio-Rad PDQuest software. The values represented in the table as x/y indicate that ‘x’ refers to the differentially expressed proteins and ‘y’ refers to the total proteins detected in each statistical test.

Figure 26: Reference reference master gels used in the PDQuest analysis of spot data

pH 4→ 7

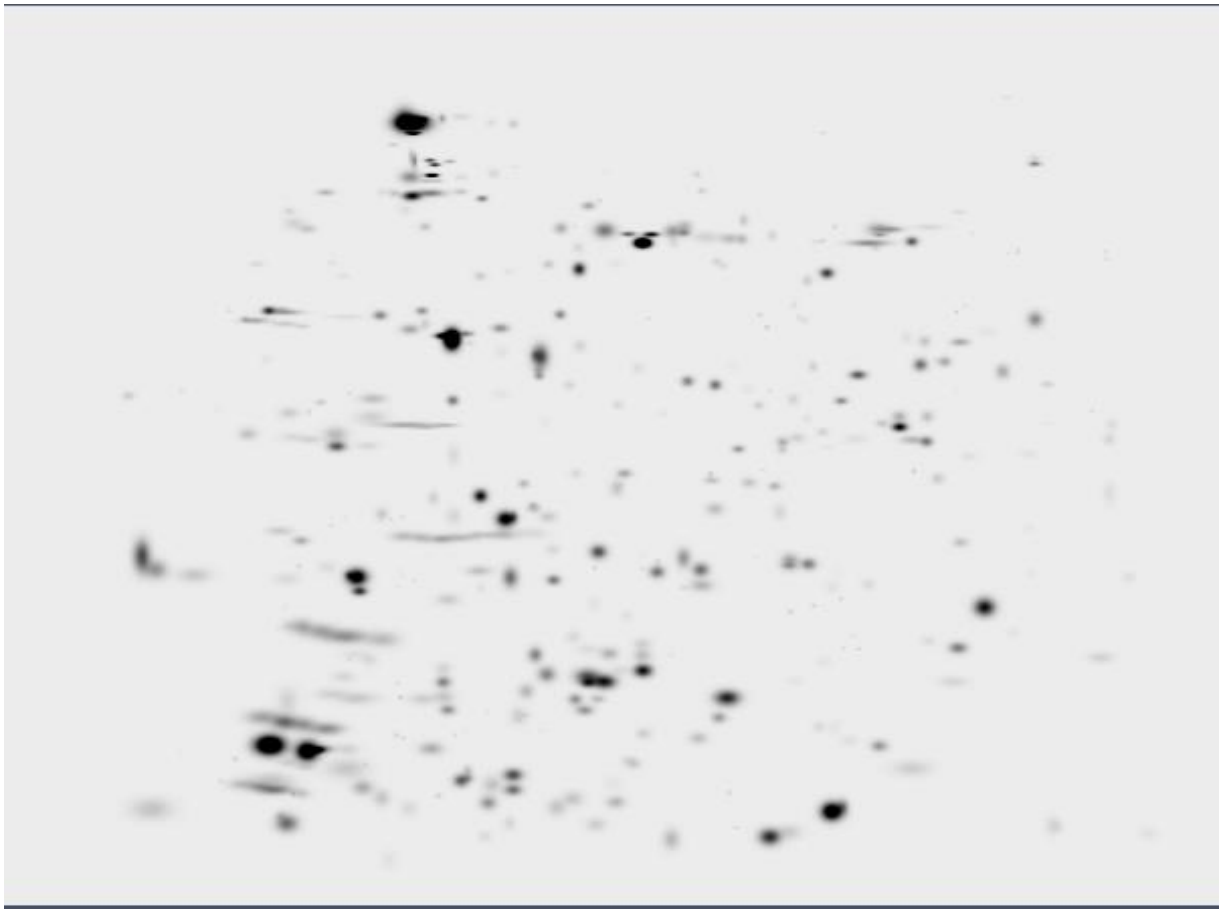


Figure 26.1: Reference master gel obtained from PDQuest analysis of *E. cancerogenus* spot data.

10.7.2 Differential expression of proteins in *E. cancerogenus*

Out of a total of 132 spots that were analysed, 98 proteins were identified using ESI LC MS/MS in *E. cancerogenus*. From the 98 proteins that were identified, 23 protein spots were differentially expressed in mucin enriched media. Seven proteins were shown to be up-regulated in both types of mucin enriched media (SSP numbers- 4008, 5306, 5601, 5607, 5701, 5702, 6203, 6605), three proteins were up-regulated in mucin Type II enriched media (SSP numbers- 5102, 7602, 8205) and 5 proteins in mucin Type III enriched media (SSP numbers- 6306, 6401, 6408, 8707). Four

proteins were found to be down-regulated in both types of mucin media (SSP numbers 3002, 5003, 7001, 8206), six proteins were found to be down-regulated in mucin Type II enriched media (SSP numbers 3601, 5302, 6303, 6408, 8707, 9601) and five proteins were down-regulated in mucin Type III enriched media (SSP numbers 2302, 5003, 5601, 6711, 8205) See Fig. 26.2 for graphical summary of differentially expressed proteins.

Proteins involved in translation namely the elongation factors Tu and greA were found to be down-regulated in both mucin enriched media but elongation factor Ts was found to be up-regulated in mucin Type II enriched media (refer to Fig. 26.2, Table 20) Proteins involved in carbohydrate metabolism namely phosphoglycerate kinase and enolase were found to be up-regulated in both mucin enriched media. Immune response proteins and toxins, ecotin was found to be down-regulated in both mucin enriched media except the surface antigen outer membrane protein X and outer membrane protein A which was shown to be up-regulated in both mucin enriched media. The amino acid biosynthesis protein domains for carbamate kinase, DHAP synthetase I and cytidylate kinase were also found to be up-regulated in Type III mucin enriched media. A DNA synthesis associated glutaredoxin related protein region was also identified as up-regulated in mucin Type II enriched media. The oxidative stress associated protein regions namely universal stress protein, ribonucleotide reductase region; pyruvate formate lyase region and autonomous glycyl radical cofactor were down-regulated in mucin Type III enriched media except the universal stress protein which was up-regulated in both mucin enriched media. The other up-regulated proteins include the protein transport associated periplasmic, translocation associated tolB proteins. These were found to be up-regulated in mucin Type III enriched media. The protein folding associated thiodisulphide oxidoreductase and protein metabolism associated acetoin reductase were also up-regulated in mucin Type II enriched media. Four hypothetical proteins were detected (similar to Yba B and Yfa Z protein, Hypothetical protein 20877, 17757) of which the first two (similar to Yba B and Yfa Z protein) were found to be up-regulated in mucin Type II enriched media and the hypothetical protein 20877 was up-regulated in mucin Type III enriched media. The hypothetical protein 17757 was found to be

down-regulated in mucin Type III enriched media. Refer to Appendix G for details on the putative identifications obtained using BLAST analysis.

Analyses performed using the Redfin and Samespots softwares confirmed the up-regulation of the cell wall, membrane and envelope biogenesis associated protein outer membrane surface antigen X and carbohydrate metabolism associated phosphoglycerate kinase in both mucin enriched media in *E. cancerogenus*. Transcription associated elongation factor Ts was found to be up-regulated in mucin Type III enriched media in all the software analyses in *E. cancerogenus*. Refer to Tables 16 and 17 for further details.

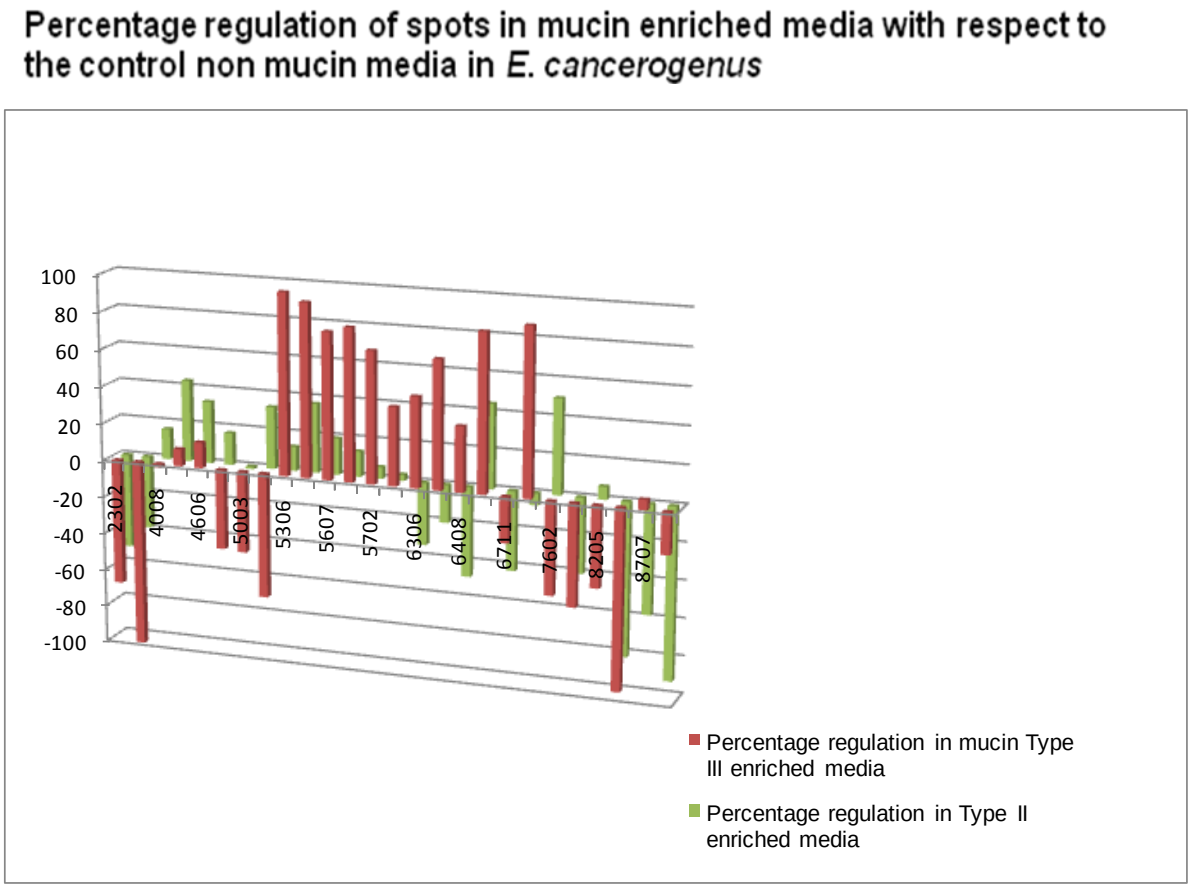


Figure 26.2: Percentage regulation of spots in mucin enriched media with respect to the control non-mucin media in *E. cancerogenus*

pH 4→ 7

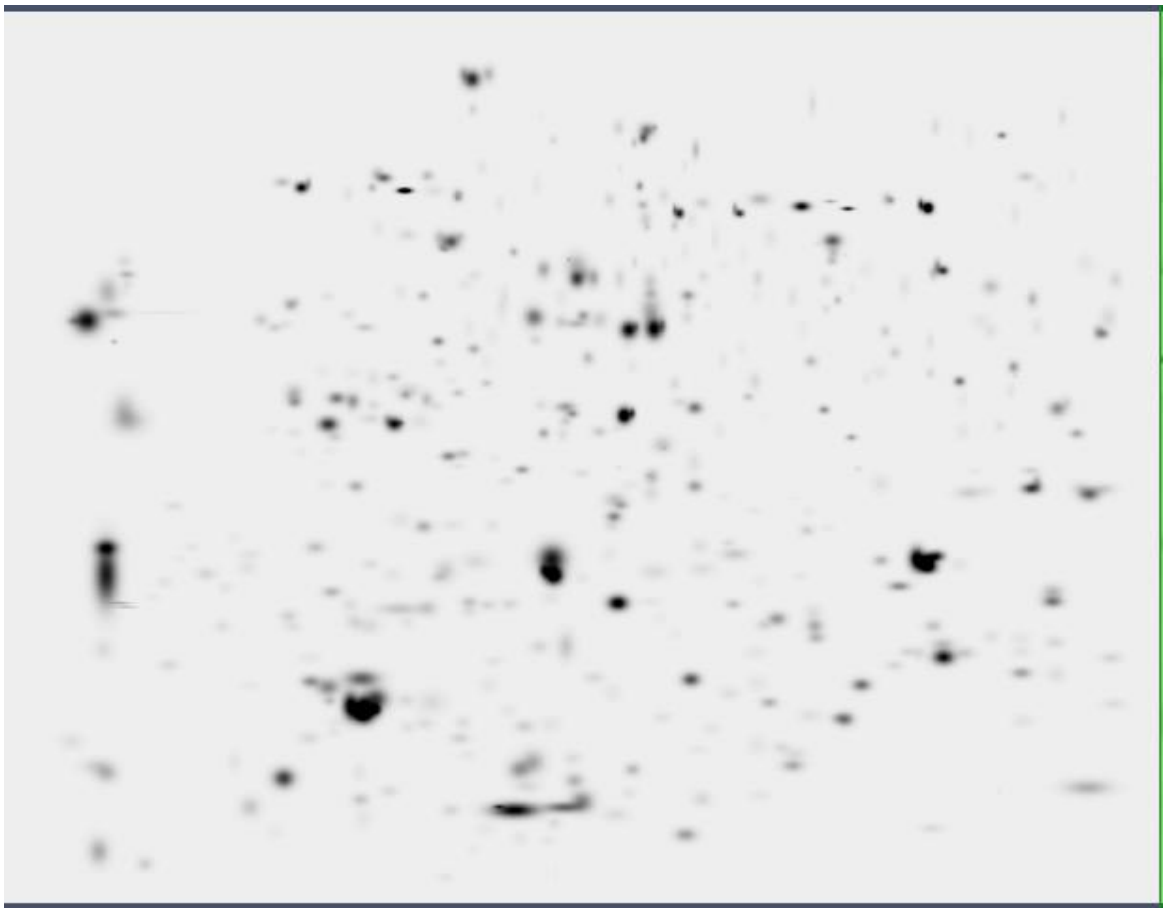


Figure 26.3: Reference master gel obtained from PDQuest analysis of *B. fragilis* spot data.

10.7.3 Differential expression of proteins in *B. fragilis*

Out of a total of 106 proteins that were analysed, 73 were identified using LC MS/MS in *B. fragilis* and 45 were found to be differentially expressed. A total of 31 spots (SSP numbers- 1104, 2101, 2102, 2401, 2701, 2704, 2804, 2806, 3303, 3404, 4502, 4601, 4607, 4608, 4702, 4804, 5106, 5201, 5205, 5402, 6102, 6103, 6202, 6206, 6301, 6304, 6802, 6807, 6901, 7301, 7401) were differentially expressed in mucin enriched media but only 5 proteins (SSP numbers- 1104, 2701, 2704, 2806 and 4804) were found to be up-regulated in mucin enriched media. SSP number 1104 was up-regulated in both types of mucin media but SSP 2701and 2704 were up-regulated in mucin Type III enriched media, absent in mucin Type II enriched media. SSP 2806 was found to be up-regulated in both mucin enriched media. Twenty-five proteins were found to be down-regulated in mucin enriched media. Eight proteins

were found to be down-regulated in both mucin enriched media (SSP numbers 1803, 2102, 3001, 3303, 4601, 5106, 5205, 6103) whereas 12 proteins were down-regulated in mucin Type II enriched media and 5 proteins were down-regulated in mucin Type III enriched media. The presence of a host immune response may trigger the up-regulation of more proteins in the bacteria (Tran *et al.*, 2003). The predominant proteins showing regulation in mucin enriched media have been indicated in Fig. 26.4 and Table 20.

The 4 protein spots that were found to be up-regulated were the hypothetical protein BF2494 (SSP 1104), proton transport and ATP synthesis associated ATP synthase subunit E (SSP 2701, 2704) and translation associated 50S ribosomal protein L7/L12 (SSP 2806).

The presence of protein isoforms may have been indicated by the identification of the same protein in more than one spot that occured adjacent to each other on the gels but not much information is available regarding the existence of isoforms in prokaryotic bacteria. Proteins that showed this phenomenon were the citric acid cycle associated malate dehydrogenase in *B. fragilis*, oxidative stress associated thiol peroxidase in *B. fragilis* and the immune response associated outer membrane proteins in *E. cancerogenus* and *B. fragilis* (refer to Figs. 30, 31, 33 and 34 for zoomed in images in the Discussion section).

The proteins that showed down-regulation included the carbohydrate metabolism associated phosphoenol pyruvate carboxylase, triose phosphate isomerase, oxidative stress associated thiol peroxidase, immune response associated putative outer membrane protein, glutathione detoxification system associated lactoylglutathione lyase, RNA protection associated putative RNA binding protein, translation associated ribosomal protein L7/L12 and elongation factor Ts, citric acid cycle associated malate dehydrogenase and chaperone associated GrpE protein. Refer to Table 21 for details. Two down-regulated hypothetical proteins were identified namely BF1203 and BF0301 and one hypothetical protein BF2494 was up-regulated. Refer to Appendix G4 for details on the putative identifications obtained using BLAST analysis. Hypothetical proteins BF2494 were found to be up-regulated in mucin Type II enriched media when analysed using both PDQuest and SameSpots softwares but not in Redfin. Similarly, BF1203 hypothetical proteins were

found to be down-regulated in both mucin Type II and III enriched media when analysed using both PDQuest and SameSpots softwares (tests were based on the Student t-test).

Analyses performed using the Redfin and Samespots softwares in *B. fragilis* confirmed the down-regulation of the carbohydrate metabolism associated triose phosphate isomerase in mucin Type II enriched media. Energy production and conversion associated malate dehydrogenase proteins were found to be down-regulated in mucin Type II enriched media but up-regulated in mucin Type III enriched media in *B. fragilis*. All analyses also showed the up-regulation of the transcription protein elongation factor Ts in mucin Type II enriched media. Student t-test was used to perform the analyses using all the three softwares where expression levels were considered based on average values in each condition.

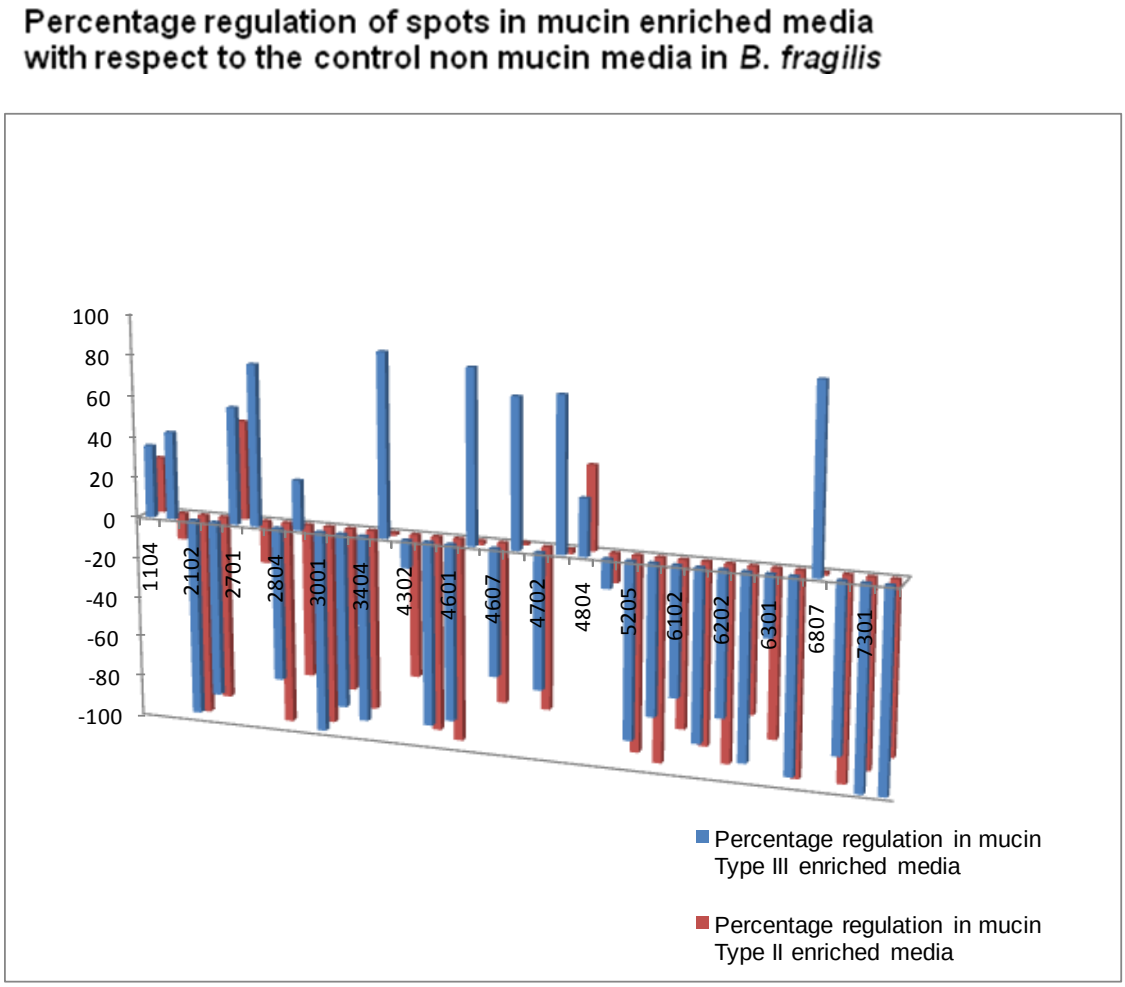


Figure 26.4: Percentage regulation of spots in mucin enriched media with respect to the control non-mucin media in *B. fragilis*

The general growth rate of both *B. fragilis* and *E. cancerogenus* were not profoundly influenced by the presence of mucin in the media especially with reference to culture density.

10.7.4 Analysis of 2DE gels using Ludesi Redfin software

Reports have been generated for the 2DE gel analysis using the Ludesi Redfin software for studying the differential expression of proteins in control vs mucin enriched media in *E. cancerogenus* and *B. fragilis*. Refer to Appendix F2 for results, including p values and expression levels of *E. cancerogenus* and *B. fragilis* proteins in non-mucin, mucin Type II and Type III enriched media. Spots selected for analysis were based on significant p values of <0.05 and a fold change of 1.5 or more. Expression profile filters were also used to detect up or down regulation of proteins.

10.7.5 Analysis of 2DE gels using Nonlinear Dynamics SameSpots software

The SameSpots analysis software from Nonlinear Dynamics was also used to study the differential expression of proteins in both bacteria. The spots of interest were selected based on the analysis of variance where significant p values of <0.05 and fold changes of 1.0 or more were considered. Refer to Appendix F3 for the analysis report of *E. cancerogenus* and *B. fragilis* spot data, respectively.

10.7.6 Comparative analysis of spot data

The spots with relevant p values from each analysis were manually compared in all the three softwares (Fig 27; Tables 12-19). The proteins highlighted in ‘bold’ font in the tables indicate differential expression in all software analyses under those specific growth conditions and statistical tests. SSP number refers to the Standard spot number which has been used to identify the protein spots using the default Bio-Rad PDQuest software. Each circle of the Venn diagram indicates the differentially expressed proteins in one particular software package and the overlap regions

indicate the proteins identified commonly in two of the software packages. The central region where all the three circles overlap shows the proteins that have been differentially expressed in all the three software packages (Fig. 27). SSP numbers have been indicated for spots that remained unidentified using LC-MS analysis.

Since SameSpots works only on the basis of average expression in each growth condition, the Mann-Whitney U test, where expression of a protein in each individual gel can be considered, could not be performed using data from SameSpots. Therefore the Student t-test was used to compare SameSpots, PDQuest and Redfin data, and the Mann-Whitney U test was used to compare PDQuest and Redfin data.

The Student t-test was used to identify proteins with a significant p value (<0.05) from all three softwares and the average values were used to compare expression. The Mann-Whitney U test was used to identify proteins with a significant p value (<0.05) from all three softwares but the expression in individual gels was considered unlike the Student t-test where the average values were compared.

Figure 27: Comparative analyses of spot data using PDQuest, Redfin and SameSpots softwares

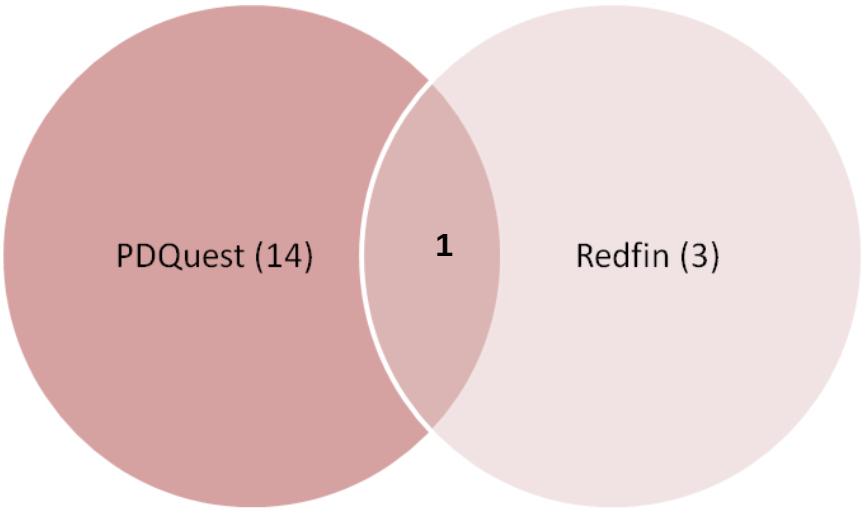


Figure 27.1: Comparative analysis of proteins differentially expressed by *E. cancerogenus* in semi-defined media without mucin vs mucin Type II enriched media by PDQuest and Redfin softwares using Mann-Whitney U test. Refer to Table 12 for details on the differentially expressed proteins in media without mucin vs mucin Type II enriched media by PDQuest and Redfin softwares using the Mann-Whitney U test.

Protein identity- *E. cancerogenus*	PDQuest analysis	Redfin analysis
SSP 7001 (Not identified)	↑	↑ (4.57, 0.0144)
SSP 6102 (Not identified)		↑ (2.55, 0.0247)
Peptidoglycan binding domain (5305)		↑ (2.51, 0.0089)
Ecotin (8206)	↓	
Glutaredoxin related protein (3002)	↓	
SSP 2802 (Not identified)	↑	
Autonomous glycyl radical cofactor (4008)	↑	
Outer membrane protein surface antigen X (6203)	↑	
Thiol disulphide oxidoreductase protein (6408)	↓	
Acetoin reductase (7602)	↑	
Tol B protein (8707)	↓	
Outer membrane protein A (4606)	↑	
Phosphoglycerate kinase (5701)	↑	
SSP 1602 (Not identified)	↑	
SSP 5302 (Not identified)	↓	
SSP 6303 (Not identified)	↓	

Table 12: Differentially expressed *E. cancerogenus* proteins in non mucin vs mucin Type II enriched media by PDQuest and Redfin softwares using the Mann-Whitney U test. ↑ indicates up-regulation of the protein, ↓ indicates down-regulation of the protein. (Fold change, p value) for differentially expressed spots has been indicated in brackets for Redfin and SameSpots softwares. SSP numbers of protein spots are indicated in the first column.

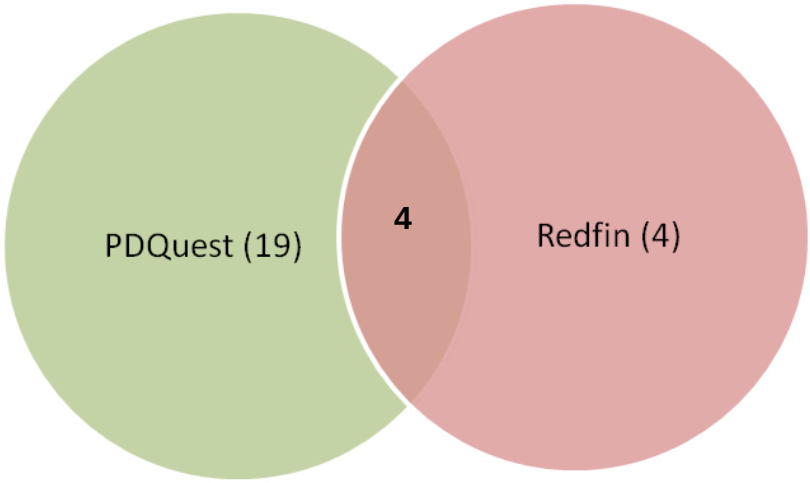


Figure 27.2: Comparative analysis of proteins differentially expressed by *E. cancerogenus* in semi-defined media without mucin vs mucin Type III enriched media by PDQuest and Redfin softwares using the Mann-Whitney U test. Refer to Table 13 for details on the differentially expressed *E. cancerogenus* proteins in media without mucin vs. mucin Type III enriched media by PDQuest and Redfin softwares using the Mann-Whitney U test.

Protein identity- *E. cancerogenus*	PDQuest analysis	Redfin analysis
Elongation factor Ts (5601)	↓	↓ (2.17, 0.0026)
Carbamate kinase (5607)	↑	↑ (2.58, 0.0296)
Periplasmic protein (5306)	↑	↑ (2.93, 0.0262)
Ecotin (8206)	↓	
Autonomous glycyl radical cofactor (4008)	↓	
Phosphoglycerate kinase (5701)	↑	
Outer membrane protein surface antigen X (6203)	↑	↑ (2.85, 0.0143)
Elongation factor greA (2302)	↓	
Glutaredoxin related protein (3002)	↓	
Hypothetical protein 17757(5003)	↓	
Universal stress protein A (5102)	↑	
Enolase (5705)	↑	
Hypothetical protein 20877 (6306)	↑	
Cytidylate kinase (6401)	↑	
2-dehydro-3-deoxy phosphooctonate aldolase (6605)	↑	
Elongation factor Tu (6711)	↓	
SSP 2402 (Not identified)	↑	
SSP 3602 (Not identified)	↑	
Tol B (8707)	↑	

Table 13: Differentially expressed *E. cancerogenus* proteins in non mucin vs mucin Type III enriched media by PDQuest and Redfin softwares using the Mann-Whitney U test

↑ indicates up-regulation of the protein

↓ indicates down-regulation of the protein. (Fold change, p value) for differentially expressed spots has been indicated in brackets for Redfin and SameSpots softwares. SSP numbers of protein spots are indicated in the first column.

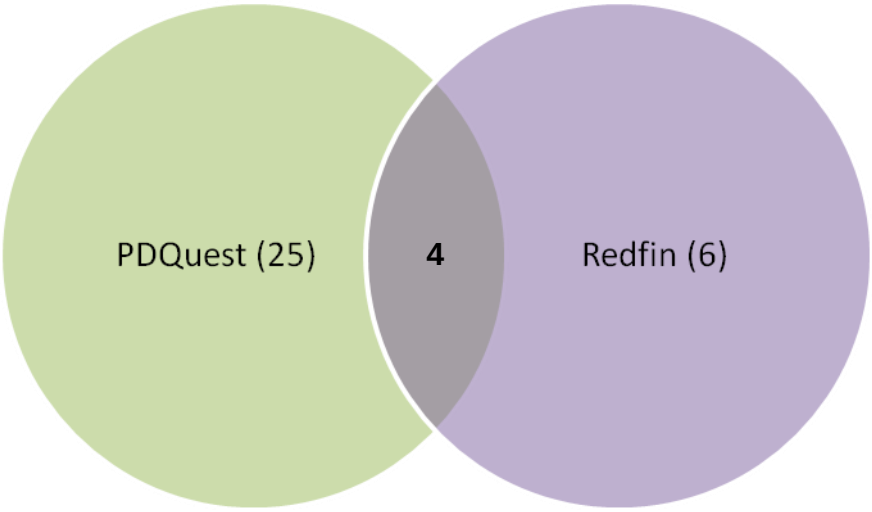


Figure 27.3: Comparative analysis of proteins differentially expressed by *B. fragilis* in semi-defined media without mucin v mucin Type II enriched media by PDQuest and Redfin software using Mann-Whitney U test. Refer to Table 14 for details on the differentially expressed B*. fragilis* proteins in media without mucin vs. mucin Type II enriched media by PDQuest and Redfin softwares using the Mann-Whitney U test based.

Protein identity- *B. fragilis*	PDQuest analysis	Redfin analysis
Putative thiol peroxidase (6304)		↓ (2.22, 0.0264)
Outer membrane protein H (6202)	↓	
Putative RNA binding protein (6102)	↓	↓ (3.66, 0.0139)
SSP 3206 (Not identified)	↓	↓ (3.11, 7.293e-4)
Putative thiol peroxidase (4601)	↓	↓ (3.07, 0.0163)
Malate dehydrogenase (4607)	↓	↓ (3.07, 3.5558e-4)
Hypothetical protein BF2494 (1104)	↑	
50S Ribosomal protein L7/L12 (2102)	↓	
Hypothetical protein BF2494 (2804)	↓	
50S Ribosomal protein L7/L12 (3001)	↓	
Triose phosphate isomerase (4502)	↓	↓ (2.69, 0.0136)
Elongation factor Ts (4702)	↓	
50S Ribosomal protein L7/L12 (5106)	↓	
Hypothetical protein BF1203 (5205)	↓	
Phosphoenol pyruvate carboxykinase (6901)	↓	
Lactoylglutathione lyase (6103)	↓	
Putative outer membrane protein H (5402)	↓	
Hypothetical protein BF 0301 (6206)	↓	
Fructose bisphosphate aldolase (6802)	↓	
Superoxide dismutase (7301)	↓	
SSP 1103 (Not identified)	↑	
SSP 1211 (Not identified)	↑	
SSP 5401(Not identified)	↓	
SSP 1107(Not identified)	↑	
SSP 1803 (Not identified)	↓	
Putative thiol peroxidase (6301)	↑	

Table 14: Differentially expressed *B. fragilis* proteins in non mucin vs mucin Type II enriched media by PDQuest and Redfin softwares using the Mann-Whitney U test. ↑ indicates up-regulation of the protein; ↓ indicates down-regulation of the protein. (Fold change, p value) for differentially expressed spots has been indicated in brackets for Redfin and SameSpots softwares. SSP numbers of protein spots are indicated in the first column.

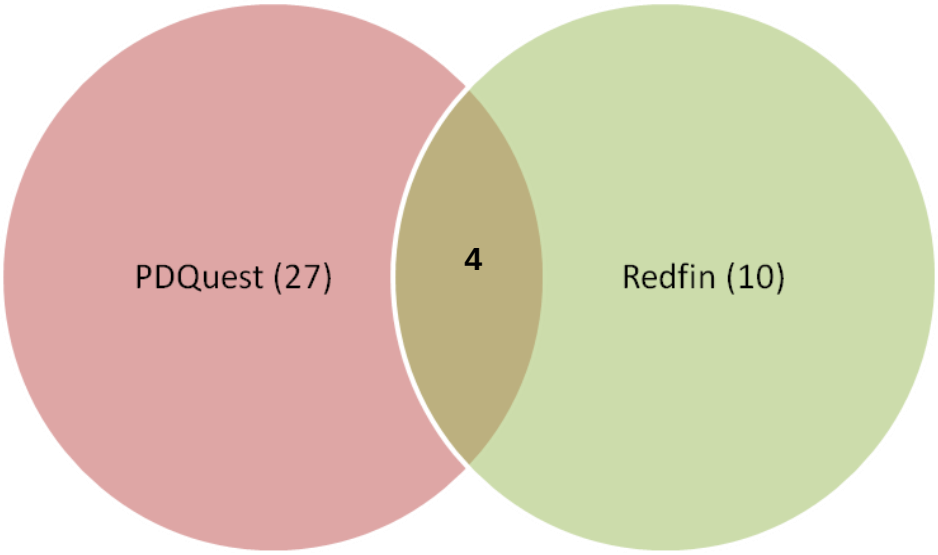


Figure 27.4: Comparative analysis of proteins differentially expressed by *B. fragilis* in semi-defined media without mucin vs mucin Type III enriched media by PDQuest and Redfin software using the Mann-Whitney U test. Refer to Table 15 for details on the differentially expressed B*. fragilis* proteins in media without mucin vs. mucin Type III enriched media by PDQuest and Redfin softwares using the Mann-Whitney U test.

Protein identity- *B. fragilis*	PDQuest analysis	Redfin analysis
Putative thiol peroxidase (6304)		↓ (2.22, 0.0264)
Outer membrane protein H (6202)	↓	↓ (4.69, 0.0176)
SSP 3206 (Not identified)		↑ (3.11, 7.293e-4)
Putative thiol peroxidase (4601)	↓	↓ (3.07, 0.0163)
SSP 1207 (Not identified)		↓ (2.65, 0.0101)
SSP 3305 (Not identified)		↓ (2.47, 0.0479)
Malate dehydrogenase (4607)	↓	↓ (2.62, 0.0211)
Malate dehydrogenase (4608)	↑	↑ (3.07, 3.5558e-4)
SSP 5002 (Not identified)		↓ (2.89, 0.0037)
50S Ribosomal protein L7/L12 (2102)	↓	
50S Ribosomal protein L7/L12 (5106)	↓	
Lactoylglutathione lyase (6103)	↓	
Triose phosphate isomerase (6302)	↓	↓ (2.69, 0.0136)
Hypothetical protein BF2494 (2804)	↓	
SSP 1103 (Not identified)	↑	
Hypothetical protein BF2494 (1104)	↑	
GrpE protein (2401)	↓	
ATP synthase subunit E (2701)	↑	
SSP 4101 (Not identified)	↑	
Malate dehydrogenase (4804)	↑	
Hypothetical protein BF1203 (5201)	↓	
Superoxide dismutase (3404)	↑	
Thioredoxin (6301)	↑	
Hypothetical protein BF1203 (5205)	↓	
Hypothetical protein BF 0301(6807)	↓	
SSP 1211 (Not identified)	↑	
SSP 1402 (Not identified)	↓	
SSP 1803 (Not identified)	↓	
SSP 2403 (Not identified)	↑	
SSP 2504 (Not identified)	↑	
ATP synthase subunit E (2704)	↑	
50S Ribosomal protein L7/L12 (2806)	↑	

Table 15: Differentially expressed *B. fragilis* proteins in non mucin vs mucin Type III enriched media by PDQuest and Redfin softwares using the Mann-Whitney U test. ↑ indicates up-regulation of the protein; ↓ indicates down-regulation of the protein. (Fold change, p value) for differentially expressed spots has been indicated in brackets for Redfin and SameSpots softwares.

SSP numbers of protein spots are indicated in the first column.

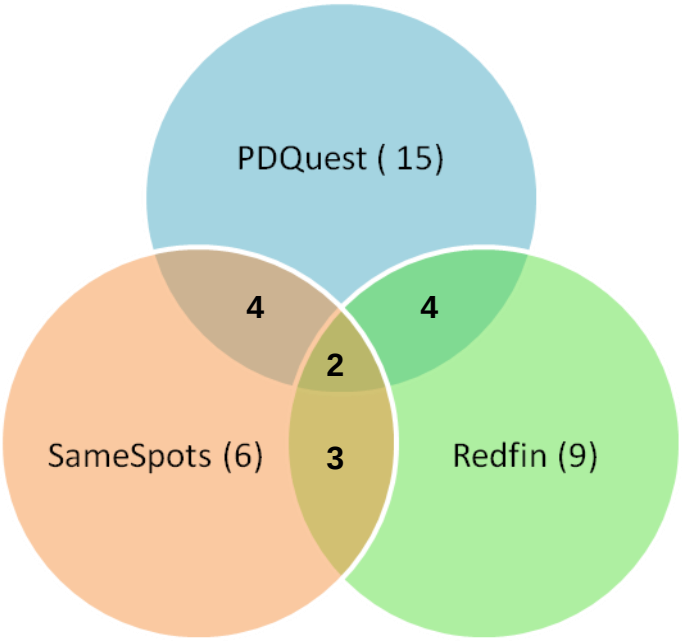


Figure 27.5: Comparative analysis of proteins differentially expressed by *E. cancerogenus in* semi-defined media without mucin vs mucin Type II enriched media by PDQuest, Redfin and Same spots software using the Student t-test. Refer to Table 16 for details on the differentially expressed *E.cancerogenus* proteins in media without mucin vs. mucin Type II enriched media by PDQuest, SameSpots and Redfin softwares using the Student t-test.

Protein identity- *E. cancerogenus*	PDQuest analysis	Redfin analysis	Samespots analysis
SSP 7001 (Not identified by mass spectrometry)	↑	↑ (4.57, 0.0144)	
SSP 6102 (Not identified)		↑ (2.55, 0.0247)	
SSP 1602 (Not identified)	↓		↓ (3.4, 0.024)
Peptidoglycan binding domain (5305)		↑ (2.51, 0.0089)	
Ecotin (8206)	↓		
Glutaredoxin related protein (3002)	↓		
Autonomous glycyl radical cofactor (4008)	↓		
Outer membrane protein X (6203)	↑	↑ (2.85, 0143)	↑ (2.3, 0.006)
Acetoin reductase (7602)	↑		
Tol B protein (8707)	↓		
Outer membrane protein II (4505)	↓		
Enolase (5705)			↑ (6.5, 0.029)
Phosphoglycerate kinase (5701)	↑	↑(2.51, 0.0089)	↑ (2.3, 0.032)
SSP 7003 (Not identified)		↑ (2.55, 0.0116)	↑ (2.3, 0.006)
Inorganic pyrophosphatase (5405)	↑	↑ (1.73, 0.0033)	
Thiol disulphide oxidoreductase protein (6408)	↓		↓ (2.2, 0.003)
SSP 2509 (Not identified)	↑		
SSP 3602 (Not identified)	↓		
SSP 5302 (Not identified)	↓		
Putative glutathione peroxidase (4701)		↑ (1.81, 0.0459)	
ABC type sugar transport system (6001)		↑ (4.41, 0.045)	

Table 16: Differentially expressed *E. cancerogenus* proteins in non mucin vs mucin Type II enriched media by PDQuest and Redfin and Samespots softwares using the Student t-test (i.e. based on averages) ↑ indicates up-regulation of the protein. ↓ indicates down-regulation of the protein. (Fold change, p value) for differentially expressed spots has been indicated in brackets for Redfin and SameSpots softwares. SSP numbers of protein spots are indicated in the first column.

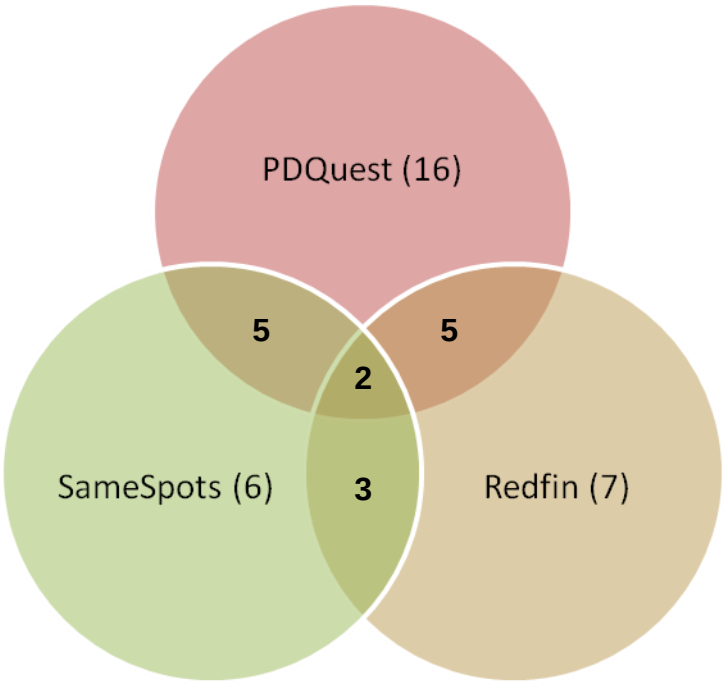


Figure 27.6: Comaparative analysis of proteins differentially expressed by *E. cancerogenus* in semi-defined media without mucin v mucin Type III enriched media by PDQuest, Redfin and SameSpots softwares using the Student t-test (i.e. based on averages). Refer to Table 17 for details on the differentially expressed *E.cancerogenus* proteins in media without mucin vs. mucin Type III enriched media by PDQuest, SameSpots and Redfin softwares using the Student t-test.

Protein identity- *E. cancerogenus*	**PDQuest analysis**	**Redfin analysis**	**Samespots software**
Elongation factor Ts (5601)	↓	↓ (2.17, 0.0026)	↓ (2.5, 0.041)
Carbamate kinase (5607)	↓	↓ (2.58, 0.0296)	
Periplasmic protein (5306)	↑	↑ (2.93, 0.0262)	
Ecotin (8206)	↓		↓ (6.0, 3.043e-004)
Autonomous glycyl radical cofactor (4008)	↓		
Phosphoglycerate kinase (5701)	↑		
Outer membrane protein surface antigen X (6203)	↑	↑ (2.85, 0.0143)	↑ (2.3, 0.006)
Elongation factor greA (2302)	↓		
Glutaredoxin related protein (3002)	↓		
Hypothetical protein 17757 (5003)	↓		↓ (2.1, 0.054)
Universal stress protein (5102)	↑		
Enolase (5705)	↑		
Outer membrane protein A (4606)	↑		
Cytidylate kinase (6401)	↑		↑ (2.5, 0.045)
DAHP synthetase (6605)	↑		
Elongation factor Tu (6711)	↑	↑ (2.55, 0.0247)	
SSP 7003 (Not identified)		↑ (2.55, 0.0116)	↑ (2.3, 0.006)
ABC type sugar transport system (6001)		↑ (4.41, 0.045)	

Table 17: Differentially expressed *E. cancerogenus* proteins in non mucin vs mucin Type III enriched media by PDQuest and Redfin and SameSpots softwares using the Student t-test (i.e. based on averages). ↑ indicates up-regulation of the protein

↓ indicates down-regulation of the protein. (Fold change, p value) for differentially expressed spots has been indicated in brackets for Redfin and SameSpots softwares. SSP numbers of protein spots are indicated in the first column.

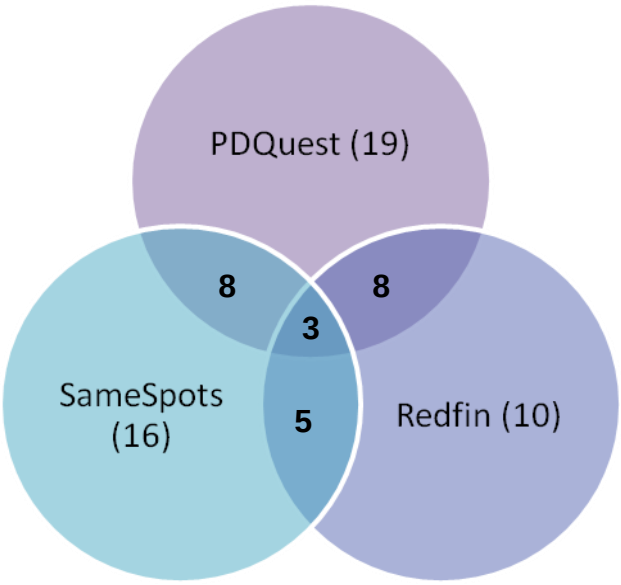


Figure 27.7: Comparative analysis of proteins differentially expressed by *B. fragilis* in semi-defined media without mucin vs mucin Type II enriched media by PDQuest, Redfin and Same Spots softwares using the Student t-test (i.e. based on averages).

Refer to Table 18 for details on the differentially expressed *B. fragilis* proteins in media without mucin vs. mucin Type II enriched media by PDQuest, SameSpots and Redfin softwares using the Student t-test.

.

Protein identity- *B. fragilis*	PDQuest analysis	Redfin analysis	Samespots software
Putative thiol peroxidase (6304)		↓ (2.22, 0.0264)	↑(4.2, 0.024)
Outer membrane protein H (6202)	↓	↓ (4.69, 0.0176)	
Putative RNA binding domain (6102)	↓	↓ (3.66, 0.0139)	
SSP 3206 (Not identified)		↓ (3.11, 7.293e-4)	↓ (1.9, 0.003)
Putative thiol peroxidase (4601)	↓	↓ (3.07, 0.0163)	
Malate dehydrogenase (4607)	↓	↓ (3.07, 3.5558e-4)	↓ (3.2, 0.002)
Hypothetical protein BF2494 (1104)	↑		↑ (5.1, 6.672e-004)
50S Ribosomal protein L7/L12 (2102)	↓		
Hypothetical protein BF2494 (2804)	↓		
50S Ribosomal protein L7/L12 (3001)	↓		↓ (2.4, 0.043)
Triose phosphate isomerase (4502)	↓	↓ (2.69, 0.0136)	↓ 4.5, 0.052)
Elongation factor Ts (4702)	↑	↑ (2.34, 0.0250)	↑ (2.4, 0.043)
50S Ribosomal protein L7/L12 (5106)	↓		
Hypothetical protein BF1203 (5205)	↓		↓ (1.5, 0.033)
Lactoylglutathione lyase (6103)	↓	↓ (2.68, 0.0211)	
Hypothetical protein 1203 (5402)	↓		
Hypothetical protein 0301 (6206)	↓		
Phosphoenol pyruvate carboxykinase (6901)	↓		
Fructose bisphosphate aldolase (6802)	↓		↓ (3.0, 0.002)
SSP 3904 (Not identified)			↓ (3.7, 0.003)
50S Ribosomal protein L7/L12 (3106)			↓ (1.7, 0.009)
Superoxide dismutase (6206)	↓		↓ (6.9, 0.004)
SSP 5806 (Not identified)			↓ (4.3, 0.028)
SSP 1101 (Not identified)			↓ (6.1, 5.017e-004)
Triose phosphate isomerase (6302)	↓	↓ (6.78, 0.0156)	
SSP 1302 (Not identified)			↓(2.0, 6.287e-004)
Adenylate kinase (4302)			↓(2.6, 0.046)

Table 18: Differentially expressed *B. fragilis* proteins in non mucin vs mucin Type II enriched media by PDQuest and Redfin and SameSpots softwares using the Student t-test (i.e. based on averages) ↑ indicates up-regulation of the protein ↓ indicates down-regulation of the protein. (Fold change, p value) for differentially expressed spots has been indicated in brackets for Redfin and SameSpots softwares. SSP numbers of protein spots are indicated in the first column.

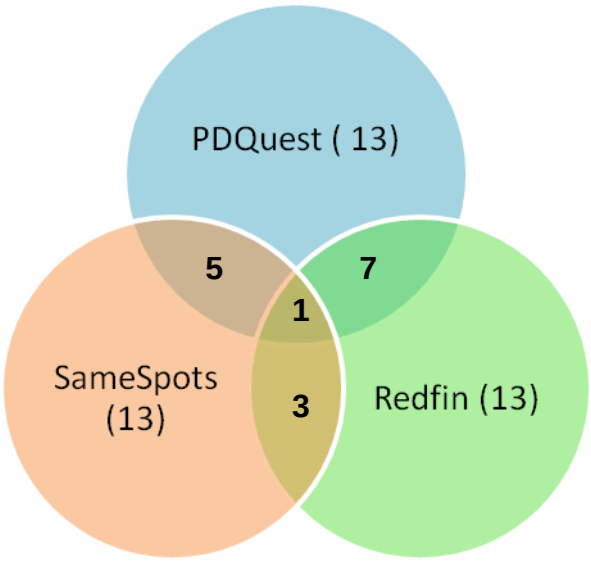


Figure 27.8: Comparative analysis of proteins differentially expressed by *B. fragilis* in semi-defined media without mucin vs mucin Type III enriched media by PDQuest, Redfin and SameSpots softwares using the Student t-test (i.e. based on averages). Refer to the Table 19 for details on the differentially expressed *B. fragilis* proteins in media without mucin vs. mucin Type III enriched media by PDQuest, SameSpots and Redfin softwares using the Student t-test.

Protein identity- *B. fragilis*	PDQuest analysis	Redfin analysis	Samespots software
Putative thiol peroxidase (6304)	↓	↓ (2.22, 0.0264)	
Outer membrane protein H (6202)	↓	↓ (4.69, 0.0176)	
SSP 3206 (Not identified)		↑ (3.11, 7.293e-4)	↑ (1.9, 0.003)
SSP 1207 (Not identified)		↓ (2.65, 0.0101)	
SSP 3305 (Not identified)		↓ (2.47, 0.0479)	
Malate dehydrogenase (4607)		↓ (3.07, 3.5558e-4)	
Malate dehydrogenase (4608)	↑	↑ (3.24, 0.0431)	↑ (1.3, 0.053)
SSP 5002 (Not identified)		↓ (2.89, 0.0037)	
50S Ribosomal protein L7/L12 (2102)	↓		↓ (2.4, 0.043)
50S Ribosomal protein L7/L12 (5106)	↓		
Lactoylglutathione lyase (6103)	↓	↓ (2.68, 0.0211)	
SSP 3904 (Not identified)			↓ (3.7, 0.003)
Adenylate kinase (4302)			↓ (2.6, 0.046)
SSP 7001 (Not identified)			↓ (3.8, 0.001)
Hypothetical protein BF2494 (1104)	↑		
Hypothetical protein BF1203 (5205)	↓		↓ (1.5, 0.033)
Hypothetical protein BF1203 (5201)	↓		↓ (2.8, 0.024)
Malate dehydrogenase (4705)			↓ (2.1, 0.047)
Putative thiol peroxidase (4701)			↑(4.2, 0.024)
SSP 3702 (Not identified)			↑ (3.6, 0.012)
SSP 1803 (Not identified)	↓		↓ (1.5, 0.049)
SSP 5103 (Not identified)		↓ (2.46, 0.0271)	↓ (1.8, 0.043)
Outer membrane protein precursor (3303)	↓	↓ (5.07, 2.848e-6)	
SSP 2403 (Not identified)	↑	↑ (2.70, 0.0381)	
GrpE protein (2401)	↓	↓ (2.66, 0.0034)	

Table 19: Differentially expressed *B. fragilis* proteins in non mucin vs mucin Type III enriched media by PDQuest and Redfin and SameSpots softwares using the Student t-test (i.e. based on averages) ↑ indicates up-regulation of the protein; ↓ indicates down-regulation of the protein. (Fold change, p value) for differentially expressed spots has been indicated in brackets for Redfin and SameSpots softwares. SSP numbers of protein spots are indicated in the first column.

10.7.7 Analysis using Redfin and SameSpots softwares

Initial analysis of results was performed using the PDQuest software using the Student t-test and hence this has been used as the default method for discussing the results. Following further analysis using the Redfin and SameSpots softwares, the differential expression data from all the three softwares were compared. PDQuest and Redfin software analyses were based on the Mann Whitney U test performed to determine the proteins showing a regulation based on the levels of expression in each individual gel in every growth condition. Analyses using PDQuest, SameSpots and Redfin softwares were based on the Student t-test performed based on the average level of expression of proteins in every growth condition.

The Mann Whitney U test based analysis of proteins showed a down-regulation of elongation factor Ts in mucin Type III enriched media and an up-regulation of the outer membrane surface antigen protein in mucin Type III enriched media in *E. cancerogenus*. Apart from these two proteins, the periplasmic proteins and carbamate kinase proteins were up-regulated in mucin Type III enriched media whereas SSP number 7001 protein which was unidentifiable from mass spectrometry was up-regulated in mucin Type II enriched media in *E. cancerogenus*. The Student t-test based analyses of proteins from all the 3 softwares showed an up-regulation of outer membrane surface antigen protein in both mucin Type II and III enriched media and an up-regulation of phoshoglycerate kinase in mucin Type II enriched media in *E. cancerogenus*. In mucin Type III enriched media in *E. cancerogenus*, the elongation factor Ts protein was found to be down-regulated.

From both the tests, it could be identified that the outer membrane surface antigen was up-regulated in mucin Type III enriched media in *E. cancerogenus* indicating that this protein may be a potential virulence factor whose expression may be associated with the presence of mucin in the growth media.

The Mann Whitney U test based analysis of proteins in *B. fragilis* showed a down-regulation of the RNA binding protein, thiol peroxidase, malate dehydrogenase and triose phosphate isomerase proteins in mucin Type II enriched media. In mucin Type III enriched media, down-regulation of the outer membrane protein H, thiol peroxidase and triose phosphate isomerase proteins was observed in *B. fragilis*.

The Student t-test based analyses of proteins from all the 3 softwares showed a down-regulation of malate dehydrogenase, triose phosphate isomerase and an up-regulation of elongation factor EFTs in mucin Type II enriched media in *B. fragilis*. In mucin Type III enriched media only malate dehydrogenase was found to be up-regulated in all the 3 softwares.

From both the tests, it could be identified that malate dehydrogenase was found to be up-regulated in mucin Type III enriched media and down-regulated in mucin Type II enriched media suggesting that they may be existing as isoforms. Triose phosphate isomerase was also found to be down-regulated in mucin Type II enriched media suggesting that these proteins may be associated with the pathogenicity of *B. fragilis* when grown in the presence of mucin.

10.8 Liquid chromatography - Mass spectrometric analysis of protein spots

A total of 132 *E. cancerogenus* protein spots were analysed using LC-MS of which 98 spots were identified and 34 spots were found to be differentially expressed, according to PDQuest software analysis using Student t-test. Of the 34 spots that showed a regulation of expression, 23 spots were identifiable and the remaining 11 were unidentifiable. For *B. fragilis*, a total of 106 protein spots were analysed of which 73 spots were identified and 45 were found to be differentially expressed. Of the 45 spots that showed differential expression, 31 spots were identifiable and the remaining 14 were unidentifiable. Figure 28 shows the differentially expressed proteins in the reference master gel for *E. cancerogenus* and Figure 29 shows the differentially expressed proteins in the reference master gel for *B. fragilis.* Refer to Appendix sections G2 and G3 for details on the peptide match scores and identities of proteins using the Mascot search for both *E. cancerogenus* and *B. fragilis*.

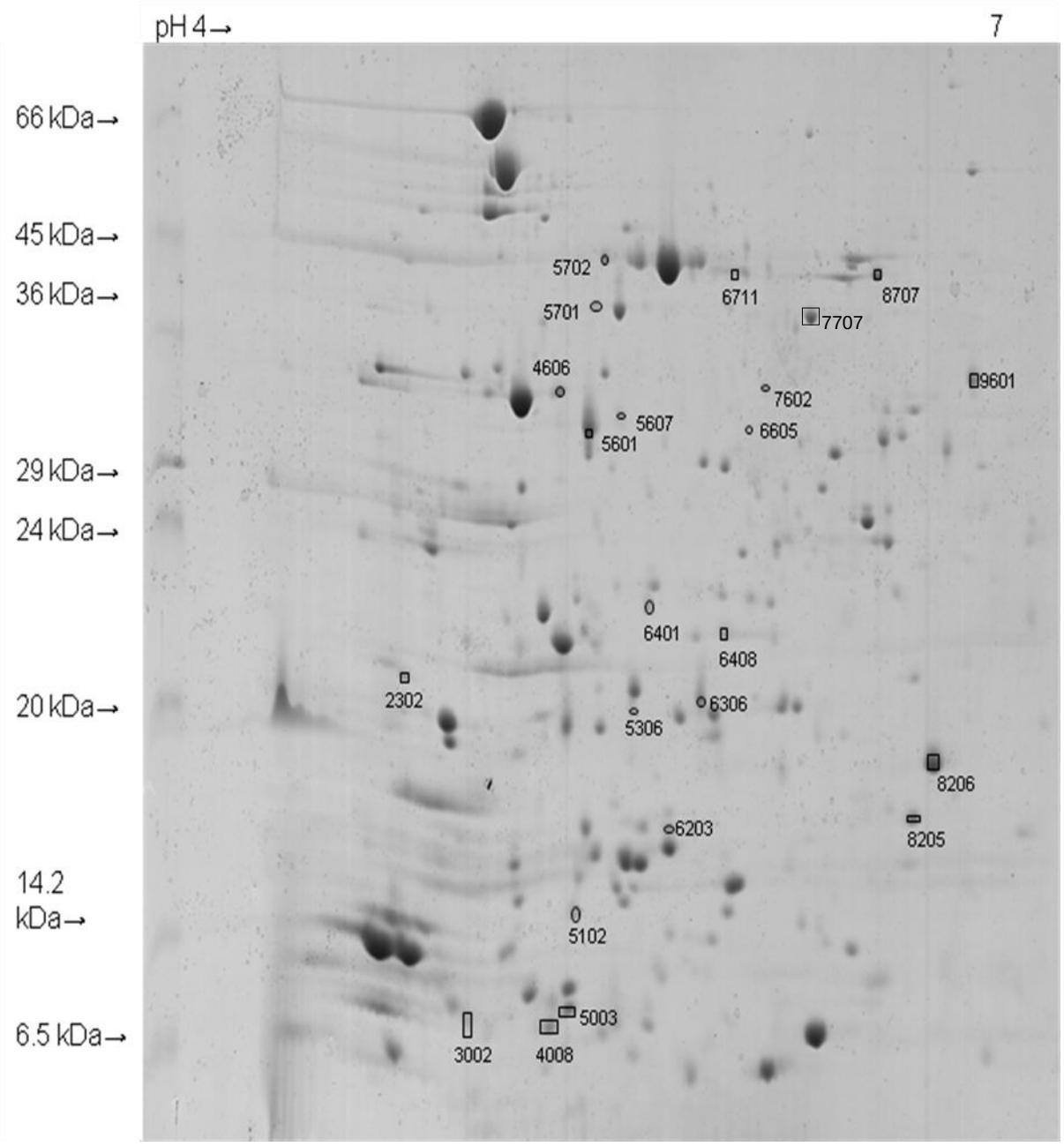


Figure 28: The SSP numbers of differentially expressed protein spots on the reference master gel of *E. cancerogenus* based on the PDQuest software analysis. ⬭ Indicates up-regulation and ▭ indicates down-regulation in mucin enriched media

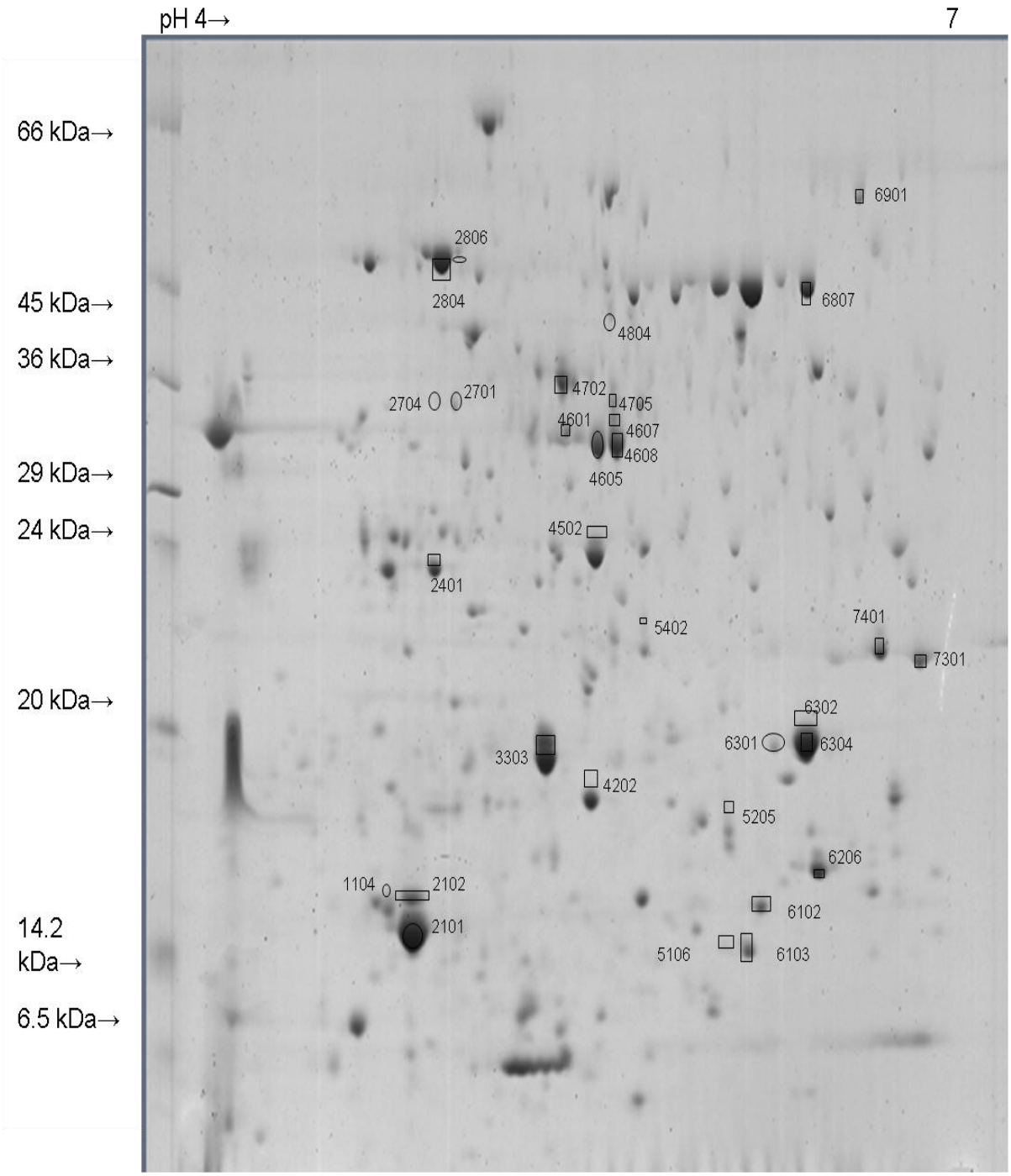


Figure 29: The SSP numbers of differentially expressed protein spots on the reference master gel of *B. fragilis* based on the PDQuest software analysis.

⬭ Indicates up-regulation and ▭ indicates down-regulation in mucin enriched media

Tables 20 and 22 represent the list of proteins identified, including the spots that were found to be differentially expressed in *E. cancerogenus* and *B. fragilis* respectively. Tables 21 and 23 represent the list of proteins identified, excluding the spots that were found to be differentially expressed in *E. cancerogenus* and *B. fragilis*.

S.No	SSP number	Identity of the protein	Function	Regulation	Peptide match score/ number of peptides	Fold change (PDQuest)	Accession number
1	2302	Elongation factor gre A	Unknown function	↓ mucin Type III media	125/3	5.9	gi|261342624
2	3002	Hypothetical protein 03947	Post translational modifications, protein turnover and chaperones	↓ mucin media	101/2	6.9	gi|261339586
3	4008	Autonomous glycyl radical cofactor GrcA	General function prediction only	↓ mucin Type III media	150/4	1.7	gi|261340878
4	4606	Omp A	Cell wall, membrane, envelope biogenesis	↑ mucin media	84/2	2.4	gi|261339284
5	5003	Hypothetical protein 17757	Protein function unknown	↓ mucin Type III media	126/2	2.2	gi|261342318
6	5102	Universal stress protein A	Signal transduction mechanisms	↑ mucin Type III media	195/7	7.2	gi|261341988
7	5306	Periplasmic protein	Function unknown	↑ mucin Type III media	130/4	5.5	gi|261339031
8	5601	Elongation factor Ts	Translation, ribosomal structure and biogenesis	↓ mucin Type III media	376/8	6.5	gi|261338832
9	5607	Carbamate kinase	Amino acid transport	↑ mucin Type III	355/6	4.9	gi|261339130

			and metabolism	media			
10	5701	Phosphoglycerate kinase	Carbohydrate transport and metabolism	↑ mucin Type II media	684/13	4.7	gi|261342323
11	5702	Phosphopyruvate hydratase	Carbohydrate transport and metabolism	↑ mucin Type III media	238/4	3.8	gi|261342208
12	6203	Outer membrane protein surface antigen X	Cell wall, membrane, envelope biogenesis	↑ mucin media	129/2	2.3	gi|261341400
13	6306	Hypothetical protein 20877	General function prediction only	↑ mucin Type III media	91/3	7.5	gi|261342932
14	6401	Cytidylate kinase	Nucleotide transport and metabolism	↑ mucin Type III media	115/3	2.4	gi|261339244
15	6408	Periplasmic protein disulphide isomerise I	Energy production and conversion	↓ mucin Type II media	189/5	6.6	gi|261342934
16	6605	2-dehydro-3-deoxyphosphooctonate aldolase	Amino acid transport and metabolism	↑ mucin Type III media	144/3	6.5	gi|261340146
17	6711	Elongation factor Tu	Translation, ribosomal structure and biogenesis	↓ mucin Type III media	246/7	3.3	gi|261341842
18	7602	Acetoin reductase	Secondary metabolites	↑ mucin Type II media	298/5	1.5	gi|261341164

			biosynthesis, transport and catabolism, signal transduction mechanisms				
19	7707	Gamma glutamyl phosphate reductase	Energy production and conversion	↓ mucin Type II media	97/2	4.7	gi|261340945
20	8205	yfaZ family protein	Unknown function	↓ mucin Type III media	195/3	1.8	gi|261339072
21	8206	Ecotin	Cell wall, membrane, envelope biogenesis	↓ mucin media	147/5	8.9	gi|261340622
22	8707	tolB translocation protein	Intracellular trafficking, secretion and vesicular transport	↓ mucin Type II media	271/7	2.6	gi|261341330
23	9601	Omp A	Cell wall, membrane, envelope biogenesis	↓ mucin Type II media	152/3	5.7	gi|261339284

Table 20: Table represents *E. cancerogenus* proteins that have been identified to be expressed in all three conditions with any up or down-regulation using PDQuest software. Regulation in mucin media in the table refers to both mucin Type II and Type III enriched media.

S.No	SSP number	Identity of the protein	Function	Regulation	Peptide match score/ number of peptides
1	1108	50S ribosomal protein L7/L12	Translation, ribosomal structure and biogenesis	None	155/5
2	1703	Porin 02457	Cell wall, membrane and envelope biogenesis	None	55/2
3	2001	Hypothetical protein 03856	Unknown function	None	116/8
4	2104	50S ribosomal protein L7/L12	Translation, ribosomal structure and biogenesis	None	134/6
5	2305	PTS system II A subunit	Carbohydrate transport and metabolism	None	237/7
6	2403	Heat shock protein grpE	Post translational modifications, protein turnover and chaperones	None	77/2
7	2506	Heat shock protein grpE	Post translational modifications, protein turnover and chaperones	None	60/2
8	3808	Trigger factor protein	Post translational modifications, protein turnover and chaperones	None	573/12
9	3809	Trigger factor protein	Post translational modifications, protein	None	71/2

			turnover and chaperones		
10	3901	Hypothetical protein 00599, Dnak chaperone protein	Heat shock proteins	None	560/9
11	3902	Hypothetical protein 00599, Dnak chaperone protein	Heat shock proteins	None	561/12
12	4101	Hypothetical protein 03459, Autonomous glycyl radical cofactor	General function prediction only	None	185/3
13	4202	Hypothetical protein 03416, 5-methylaminomethyl-2-thiouridylate methyltransferase	Translation, ribosomal structure and biogenesis	None	126/3
14	4401	Peroxiredoxin 01092	Post translational modifications, protein turnover and chaperones	None	191/4
15	4402	00804 inorganic pyrophosphatase	Energy production and conversion	None	71/2
16	4505	Omp II, 01735	Cell wall, membrane, envelope biogenesis	None	278/5
17	4602	Omp A	Cell wall, membrane, envelope biogenesis	None	468/16
18	4701	Putative glutathione peroxidase	Post translational modification, protein turnover,	None	137/4

			chaperones		
19	4812	Porin 02457	Cell envelope biogenesis	None	84/2
20	4814	Tol C protein	Intracellular trafficking, secretion and vesicular transport	None	175/4
21	4901	Molecular chaperone DnaK	Post translational modification, protein turnover, chaperones	None	60/2
22	5002	01181, YbaB family, hypothetical protein	Protein function unknown	None	166/3
23	5107	Riboflavin synthase subunit β, 01103 hypothetical protein, 6,7, dimethylribityllum azine synthase	Biosynthesis of riboflavin	None	265/4
24	5207	DNA binding protein, 02636	General function prediction only	None	218/8
25	5305	Peptidoglycan binding domain, 01554	Cell wall degradation	None	111/3
26	5401	Inorganic pyrophosphatase, 00804	Energy production and conversion	None	160/5
27	5403	Inorganic pyrophosphatase, 00804	Energy production and conversion	None	185/4
28	5405	Inorganic pyrophosphatase	Energy production and conversion	None	235/6
29	5406	Inorganic	Energy production and	None	163/5

		pyrophosphatase	conversion		
30	5604	Elongation factor Ts, 00441	Translation, ribosomal structure and biogenesis	None	259/5
31	5704	Phosphoglycerate kinase	Carbohydrate transport and metabolism	None	638/15
32	5705	Enolase, 04083	Carbohydrate transport and metabolism	None	483/14
33	6001	ABC type sugar transport system	Carbohydrate transport and metabolism	None	85/2
34	6202	OmpX, 01529	Cell wall, membrane, envelope biogenesis	None	247/3
35	6307	F0F1 ATP synthase subunit B	Energy production and conversion	None	186/4
36	6603	01449, hypothetical protein (DNA uptake, outer membrane assembly)	Function unknown	None	328/6
37	6702	Phosphopyruvate hydratase, 04083	Carbohydrate transport and metabolism	None	479/9
38	7301	Single stranded DNA binding protein	General function prediction only	None	219/4
39	7401	03228, hypothetical protein, TRX like ferredoxin family, NADH: ubiquinone oxidoreductase subunit E	Energy production and conversion	None	74/1

40	7403	02508, hypothetical protein, phage shock protein psp A	Lipid transport and metabolism	None	76/2
41	7501	Hypothetical protein, 2,5 diketo D gluconate reductase A	Biosynthesis of vitamin C, carbohydrate metabolism	None	152/4
42	7704	01034, Aldehyde dehydrogenase	Energy production and conversion	None	87/2
43	7710	03735, Flavodoxin	Energy production and conversion	None	247/7
44	8602	04209, protein of unknown function	Unknown function	None	126/3
45	8604	Phosphopyruvate hydratase	Carbohydrate transport and metabolism	None	635/16
46	8610	01936, Aldose-1-epimerase	Carbohydrate transport and metabolism	None	115/2
47	8611	03358, Dihydrodipicolinate synthase	Amino acid transport and metabolism	None	78/2
48	8704	00565, Peptidyl-prolyl cis trans isomerase SurA	Post translational modification, protein turnover and chaperones	None	223/4

Table 21: Table represents *E. cancerogenus* proteins that have been identified to be expressed in all three conditions without regulation using PDQuest software

S. No	SSP number	Identity of the protein	Function	Regulation	Peptide match score/ number of peptides	Fold change (PDQuest)	Accession number
1	1104	BF2494, Hypothetical protein	Mediates protein-protein interactions	↑ mucin media	366/8	3.1	gi|60681974
2	2101	50S ribosomal protein L7/L12	Translation, ribosomal structure and biogenesis	↓ mucin Type II media	265/4	1.5	gi|60683451
3	2102	50S ribosomal protein L7/L12	Translation, ribosomal structure and biogenesis	↓ mucin media	347/30	9.9	gi|60683451
4	2401	GrpE protein	Post translational modification, protein turnover and chaperones	↓ mucin Type III media	83/2	8.9	gi|60681311
5	2701	ATP synthase subunit E	Energy production and conversion	↑ mucin Type III media	197/4	5.3	gi|60682222
6	2704	ATP synthase subunit E	Energy production and	↑mucin Type III media	175/5	2.9	gi|60682222

			conversion				
7	2804	TRP repeat containing protein, 2494	Mediates protein-protein interactions	↓ mucin Type II media	393/7,366/8	8.8	gi|60681974
8	2806	50S ribosomal protein L7/L12	Translation, ribosomal structure and biogenesis	↑mucin Type III media	165/6	2.6	gi|60683451
9	3303	Omp precursor	Cell wall, membrane, envelope biogenesis	↓ mucin media	184/8	8.3	gi|60680025
10	4202	Omp precursor	Cell wall, membrane, envelope biogenesis	↓ mucin media	473/13	4.4	gi|60680025
11	4502	Triose phosphate isomerase	Carbohydrate transport and metabolism	↓ mucin Type II media	183/4	9.3	gi|60683174
12	4601	Putative thiol peroxidase	Post translational modification, protein turnover and chaperones	↓mucin media	112/2	9.3	gi|60682249

13	4605	Thioredoxin	Post translational modification, protein turnover and chaperones	↓mucin Type II media	140/5	4.2	gi|60680203
14	4607	Malate dehydrogenase	Amino acid transport and metabolism	↓ mucin Type II media	159/4	7	gi|60683199
15	4608	Malate dehydrogenase	Amino acid transport and metabolism	↑mucin Type III media	85/2	3.7	gi|60683199
16	4702	Elongation factor Ts	Translation, ribosomal structure and biogenesis	↓ mucin Type II media	331/7	7.3	gi|60683229
17	4705	Malate dehydrogenase	Amino acid transport and metabolism	↓mucin media	159/4	3.7	gi|60683199
18	4804	Malate dehydrogenase	Amino acid transport and metabolism	↑mucin Type III media	171/5	3.5	gi|60683199
19	5106	50S ribosomal protein L7/L12, Hypothetical	Translation, ribosomal structure and	↓ mucin media	93/1, 60/3	1.4	gi|60683451

		protein 2161	biogenesis				
20	5205	Hypothetical protein 1203, putative anti-sigma factor	Signal transduction mechanisms	↓mucin media	403/7	9.1	gi|60680695
21	5402	Hypothetical protein 1203	Signal transduction mechanisms	↓ mucin Type II media	361/9	8.7	gi|60680695
22	6102	Putative RNA binding protein	General function prediction only	↓ mucin Type II media	159/4	7.3	gi|60682207
23	6103	Lactoylglutathione lyase	Amino acid transport and metabolism	↓mucin media	195/3	8.7	gi|60682596
24	6206	Superoxide dismutase (Fe), hypothetical protein 0301	General function prediction only	↓mucin Type II media	100/3, 136/2	8.1	gi|60679842
25	6301	Thioredoxin	Post translational modification, protein turnover, chaperones	↑ in mucin Type III media	87/2	5.6	gi|60680203
26	6302	Triose phosphate isomerase	Carbohydrate transport and	↓ mucin media	87/2	9.8	gi|60683174, gi|60682249

			metabolism				
27	6802	Fructose-bisphosphate aldolase	Carbohydrate transport and metabolism	↓mucin media	196/6	8.3	gi|60682607
28	6807	Conserved hypothetical protein	Unknown function	↓ mucin media	610/14	4.1	gi|60491969
29	6901	Phosphoenol pyruvate carboxykinase	Carbohydrate transport and metabolism.	↓ mucin Type II media	663/14	9.1	gi|53715725
30	7301	FeS Superoxide dismutase	Inorganic ion transport and metabolism	↓ mucin media	466/37	9.6	gi|60682036
31	7401	Ribosome recycling factor	Translation, ribosomal structure and biogenesis	↓mucin media	235/5	9.2	gi|60680182

Table 22: *B. fragilis* (Table represents proteins that have been identified to be expressed in all three conditions with any up or down regulation using PDQuest software). Regulation in mucin media in the table refers to both mucin Type II and Type III enriched media.

S. No	SSP number	Identity of the protein	Function	Regulation	Peptide match score/ number of peptides
1	0605	BF2494, TRP repeat containing protein similar to tetratricopeptide repeat family	Mediates protein-protein interactions	None	45/2
2	1208	50S ribosomal protein L7/L12	Translation, ribosomal structure and biogenesis	None	141/4
3	2106	50S ribosomal protein L7/L12	Translation, ribosomal structure and biogenesis	None	121/2
4	3001	50S ribosomal protein L7/L12	Translation, ribosomal structure and biogenesis	None	93/1
5	3105	50S ribosomal protein L7/L12	Translation, ribosomal structure and biogenesis	None	160/2
6	3302	Co-chaperonin GroES	Post translational modification, protein turnover and chaperones	None	48/1
7	3401	FeS superoxide dismutase	Inorganic ion transport and metabolism	None	211/5
8	3403	FeS superoxide dismutase	Inorganic ion transport and metabolism	None	466/38
9	3902	Chaperone dnaK	Post translational modification,	None	345/10

			protein turnover and chaperones		
10	4302	Adenylate kinase	Nucleotide transport and metabolism	None	105/2
11	4408	GrpE protein	Post translational modification, protein turnover and chaperones	None	98/3
12	4701	Putative thiol peroxidase	Post translational modification, protein turnover and chaperones	None	172/5
13	5201	Hypothetical protein 1203, putative anti-sigma factor	Signal transduction mechanisms	None	133/3
14	5204	Hypothetical protein 1203	Signal transduction mechanisms	None	361/9
15	5808	Enolase	Carbohydrate transport and metabolism	None	469/8
16	5810	Phosphopyruvate hydratase	Carbohydrate transport and metabolism	None	354/6
17	6202	Putative OmpH	Cell wall, membrane, envelope biogenesis	None	60/1
18	6304	Thiol peroxidase	Carbohydrate transport and metabolism	None	484/18
19	6709	Hypothetical protein 1004	Protein function unknown	None	358/9
20	6801	Chaperonin groEL	Post translational	None	703/12

			modification, protein turnover, chaperones		
21	6803	Enolase	Carbohydrate transport and metabolism	None	179/4

Table 23: *B. fragilis* (Table represents proteins that have been identified to be expressed in all three conditions without any up or down regulation using PDQuest software)

MASCOT search results with the peptide scores and identity of the protein spots can be seen in Appendix G.

This study was able to identify 62% of the differentially expressed proteins using LC-MS analysis in *E. cancerogenus* and 67% of the differentially expressed proteins were identified in *B. fragilis*.

10.9 Mass spectrometric analysis of proteins

Formic acid (FA) was used in the preparation of buffers for LC-MS analysis instead of trifluoroacetic acid (TFA). This is because TFA has known negative effects which include suppression of sensitivity and ionisation of protein samples by binding and pairing to the basic groups of peptides when ionised in the gaseous phase. At a concentration of 0.1% (w/v) or more, TFA is capable of preventing spray formation or nebulisation due to the generation of a high surface tension during the ionisation process (Anon, 2002). Also, better resolution of peaks has been observed during the separation of proteins using liquid chromatography (Anon, 2002). Hence this buffer was used in our study.

Sample carry over was another potential problem that was found to influence the results obtained from the MS data. This was overcome by performing water runs in between the sample runs to remove any of the remaining peptides present in the monolith column.

Concentration of protein present in a spot was another important issue for determining its identity. Determining the concentration of the protein to be loaded to identify it from the LC-MS analysis was essential since the differential protein expression profile contained several different proteins of varying intensities. Since at least six replicates of each gel condition was generated, each of the spot was processed separately and pooled together for analysis. Initial trial experiments were performed by cutting out six sets of faint, medium and dark spots. It was observed that three or four faint spots pooled together gave a positive identification for the majority protein on MS analysis. On average, 4 spots were pooled together to detect fainter spots whereas 2 or 3 replicate spot samples were pooled together to detect dark and medium spots respectively.

Some of the spots remained undetectable because of the very low concentrations or due to sample loss during processing. Pressure variations in LC set up and needle blockages in the MS set up resulted in loss of some of the samples too. The autosampler that transferred the sample on to the column used an injection needle to pick up the sample from the multi-well microtitre plate and care had to be taken to avoid the pipetting of air bubbles along with the sample into the wells since this sometimes resulted in the blockage of the system. Frequent ultra-sonication and cleaning of the needles and the shield in the MS system helped to prevent blockages.

The possibility of adjacent spots overlapping each other on the gel suggests that the cutting out of spots need to be very accurate. It was also a possibility that some of the adjacent overlapping spots were isoforms resolving poorly. Some of the adjacent overlapping spots were identified to be the same protein. The inability to obtain a positive identity for some protein spots may be attributed to the fact that proteins were lost during processing, freeze drying or experimental errors. Several steps involved the solubilisation, freeze drying and resuspension of proteins which added to the loss of proteins.

In order to avoid random matching of measured masses to peptides in the MASCOT sequence database, the searches were restricted to the grouping of 'Eubacteria.' Random matching results could also provide false positive results and hence certain criteria were set up to segregate false positives from correct identifications. This included selection of results where the protein score was more than 40 and at least 2 peptides matched to the query sequence.

A total of 23 differentially expressed proteins were identified out of 34 in *E. cancerogenus* and 31 proteins were identified out of 45 in *B. fragilis*. The default statistical analysis used in PDQuest was the Student t-test. More than 60% of the differentially expressed proteins were identified in both cases but a substantial amount of the protein samples were lost due to processing or experimental errors.

The LC-MS spectra generated from the protein spots were searched through the MASCOT server in the NCBI and SWISS PROT databases. It was found that the NCBI prioritised the *B. fragilis* strain YCH76 over the NCTC 9343 strain. The exact reason for this is unknown. The significant hits in the YCH76 strain prompted us to

perform NCBI blast searches with its NCTC 9343 counterparts and it was observed that the matched peptide sequences were identical in both cases. There is a possibility that the unique classification, categorisation and annotation of sequences in the NCBI database is responsible for the results obtained. Similar searches in the SWISS PROT database produced matching peptide sequences in the NCTC 9343 strain but the ion scores were found to be lower. This may be because SWISS PROT database has a manually annotated protein sequence record (Mizrachi, 2002). Previous work by other scientists have suggested the presence of similar problems in data analysis (Ingrell, 2009). Refer to the second section of the discussion for further details.

Differential expression of virulence associated proteins like neuraminidases, sialidases, mucinases and glycoside hydrolases were expected to be evident from the proteomic experiments but surprisingly, none of these proteins were identified. B. fragilis, known to exhibit a well established catabolic pathway in its activity against sialic acids by producing silaidases did not express the protein in the presence of mucin, unless it was expressed in undetectable amounts. The reason behind repression of these enzymes may be due to the presence of glucose in the semi-defined growth media in addition to mucin which may be inducing a catabolite repression.

It may be interesting to perform specific assays to determine the presence of these enzymes in the cell free extracts since they have not been detected from the proteomics experiments.

10.10 Expression of proteins involved in translation, ribosomal structure and biogenesis

The ribosome recycling factor (RRF) protein was found to be down-regulated in mucin enriched media in *B. fragilis* but no regulation was observed in *E. cancerogenus*. The protein plays a vital role in the recycling of ribosomes after completion of protein synthesis. The two main signal transduction associated proteins required for the release of ribosomes from the mRNA are elongation factor (EF-G) and the ribosome recycling factor (originally known as the ribosome recycling factor (RRF); (Wilson *et al.*, 2005). RRF dissociates ribosomes from mRNA after termination of translation and is essential for bacterial growth (Janosi *et al*., 1994). The activity of this protein is important for the availability of free ribosomes for further continuation of protein synthesis. These proteins play an important role in inducing immune responses within the host during infection by activating the production of IgG antibodies (Cassataro *et al.*, 2007). This suggests that these proteins may be up-regulated during adverse growth conditions but has been found to be down-regulated in mucin enriched media in *B. fragilis*. Previous studies have shown that the growth of *Streptococcus oralis* in acidic conditions resulted in an up-regulation of these surface-associated proteins (Wilkins *et al*., 2003).

The 50S ribosomal protein L7/12 was found to be down-regulated in mucin enriched media in *B. fragilis* even though no regulation was observed in *E. cancerogenus*. The main functions of the 50S ribosomal subunit L7/L12 is mRNA directed protein synthesis, increase in polypeptide synthesis and reduction of missense error rate. Variable expression levels of 50S ribosomal subunit L7/L12 were observed in mucin media in *B. fragilis* which could be related to the possible existence of isoforms even though this phenomenon has not been observed in bacteria. But a general down-regulation of the protein was observed in mucin enriched media except SSP number 2806 which shows an up-regulation in expression.

10.11 Expression of proteins involved in transcription

Elongation factor Ts was found to be down-regulated in mucin enriched media in *B. fragilis* and *E. cancerogenus*. Interestingly, it has been known from previous studies in *S. pneumoniae* that opaque invasive variants of the bacteria were found to exhibit a down regulation of elongation factor Ts when compared to the less invasive variants that expressed higher amounts of the elongation factor (Overweg *et al.*, 2000). The elongation factor Ts plays an important role in protein synthesis by acting as a nucleotide exchange factor that is required for the regeneration of elongation factor Tu (EF-Tu) from its inactive EF-Tu-GDP form to the active form EF-Tu-GTP. This enables EF-Tu to interact with the next incoming amino acid (Hwang & Miller, 1985).

Elongation factor EF-Tu was found to be down-regulated in mucin enriched media in *E. cancerogenus* even though the protein showed no differential expression in *B. fragilis*. EF-Tu is an intracellular protein that contains a number of other proteins associated with it and is involved in translation elongation factor activity and GTP binding. It plays an important role in the translation process by initiating the selection and binding of the cognate amino-acyl t-RNA to the acceptor A site (Pape *et al.*, 1998). They may be associated with the translation and biogenesis of ribosomal structures apart from being involved in transcription. Differential expression studies in mycobacteria showed the down-regulation of these proteins until they came in contact with the host macrophages which indicates that the expression of these proteins may be triggered by the presence of a host immune response (Monahan *et al.*, 2001).

Putative RNA binding protein domains were down-regulated in mucin enriched media in *B. fragilis*. These proteins play an important role in regulating transcription termination; protect the mRNA from degradation during unfavourable growth conditions like cold shock and nucleic acid recognition (Stulke, 2002). They are capable of stopping translation by preventing the binding of mRNA to ribosomes and can control the secondary structure formation of mRNA by promoting or preventing

translation initiation thereby modulating metabolic processes. They induce antitermination of carbon and nitrogen catabolic genes to suit the survival of bacteria in the growth media and the CRM domain forms the conserved region of RNA binding protein with 100 amino acids involved in nucleic acid recognition (Barkan *et al.*, 2007). RNA binding attenuation protein trp, was found to be associated with the production of a putative efflux protein in *B. subtilis* suggesting that it may be a potential virulence factor (Yakhnin *et al*., 2006). Down-regulation of these proteins in mucin enriched media suggests that the adaptability of *B. fragilis* in this media is much better when compared to the control semi-defined media.

10.12 Expression of proteins involved in secondary metabolite biosynthesis, transport and catabolism

Acetoin reductase, a very essential enzyme involved in acetoin catabolism and acetoin dehydrogenation activity has been found to be up-regulated in media enriched with mucin Type II but down-regulated in media enriched with mucin Type III in *E. cancerogenus*. This protein was not identified in *B. fragilis*. The enzyme catalyses the conversion of acetoin and NADH to 2, 3 butanediol and NAD (+) which is a characteristic of anaerobic fermentation processes. This is also used in the commercial production of 2,3 butanediol (Xiao & Xu, 2007). A study in *B. subtilis* has revealed that this cytoplasmic protein is expressed under anaerobic conditions when the cells enter the late log phase where the production of butanediol can be observed whereas the early log phase shows accumulation of acetoin (Nicholson, 2008). The production of acetoin is essential for maintaining the pH of the media where it compensates for the acidic end products that are produced when glucose is present as a source of carbon in the media and this has been observed in *V. cholerae* (Kovacikova *et al*., 2005). This suggests that the growth stage of the bacteria during media harvest play a very important role in their protein expression.

10.13 Inorganic ion transport and metabolism associated protein expression

The enzyme superoxide dismutase has been identified to be down-regulated in mucin enriched media in *B. fragilis*. This protein has not been identified in *E. cancerogenus*. They are housekeeping enzymes that respond or confer resistance to oxidative stress in bacteria. Their location is periplasmic and catalyses the dismutation of superoxide into oxygen and hydrogen peroxide thus functioning as an antioxidant in cells exposed to oxygen. They are produced to allow the host to outcompete any aerobic flora or reduce reactive oxygen species. They destroy radicals that are commonly produced within cells and are toxic to biological systems. They occur as homodimers or homotetramers (Messerschmidt *et al.*, 2001). Previous studies have shown that the expressions of these enzymes are encoded by the *sodA* and *sodB* genes. The *sodA* gene was activated when the growth conditions shifted from anaerobic to aerobic conditions and the *sodB* gene was switched off under aerobic conditions (Matsumura et al., 1993). Growth of bacteria in media containing glucose and phosphate has been known to induce autoxidation (Carlsson *et al*., 1978) but a down-regulation of the enzyme results in a reduction in control of autoxidation and formation of superoxide radicals indicating that *B. fragilis* has a poorer survival rate in the presence of mucin.

10.14 Expression of proteins involved in cell wall, membrane and envelope biogenesis

Several outer membrane proteins have been differentially expressed in mucin enriched media with both *B. fragilis* and *E. cancerogenus*. Outer membrane protein A was up-regulated while outer membrane protein II was found to be down-regulated in *E. cancerogenus*. The outer membrane surface antigen X was found to be up-regulated in both mucin Type II and III enriched media in *E. cancerogenus*. In *B. fragilis*, two different spots were found to contain the same outer membrane protein precursors which were down-regulated in both cases in mucin enriched media. This could possibly indicate that the protein is present as isoforms in *B. fragilis* but no information is available in literature regarding the existence of isoforms in outer membrane proteins of bacteria. Refer to Figs. 30, 31 and 32 which indicate spots that expressed the same protein in *B. fragilis* and *E. cancerogenus*.

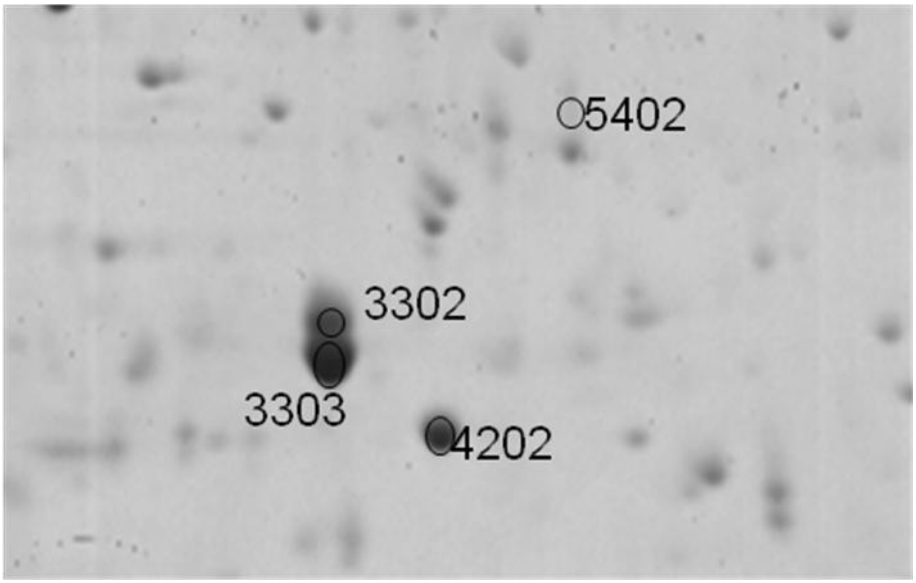


Figure 30: Figure showing the differentially expressed spots of outer membrane proteins in *B. fragilis.* SSP numbers 3302 and 5402 were identified to be outer membrane protein H and SSP numbers 3303 and 4202 were identified to be outer membrane protein precursors.

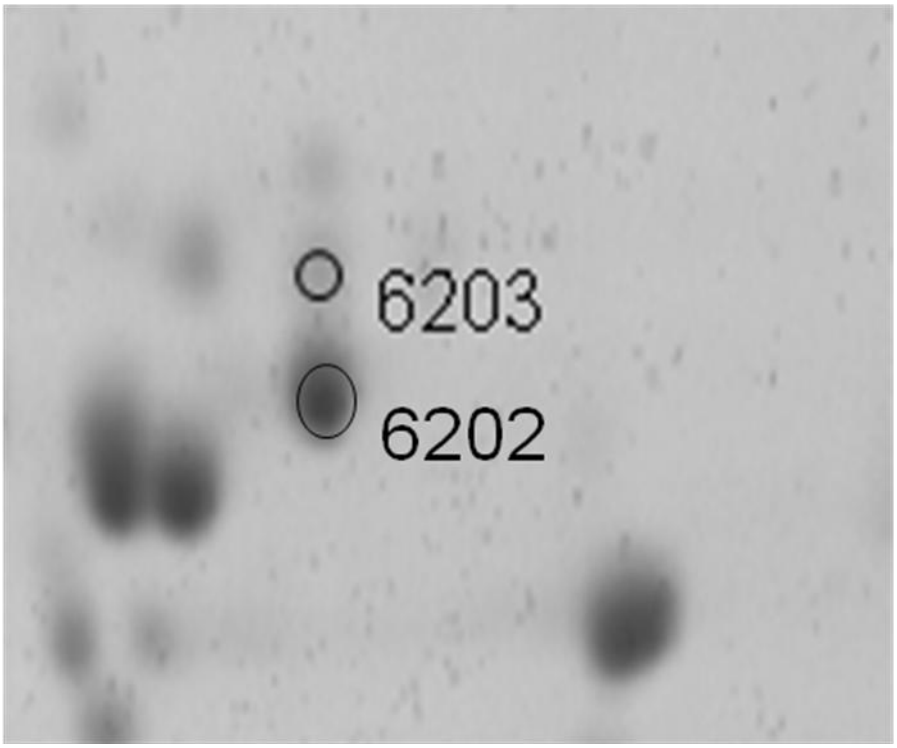


Figure 31: Figure showing the differentially expressed spots of outer membrane protein H in *B. Fragilis*

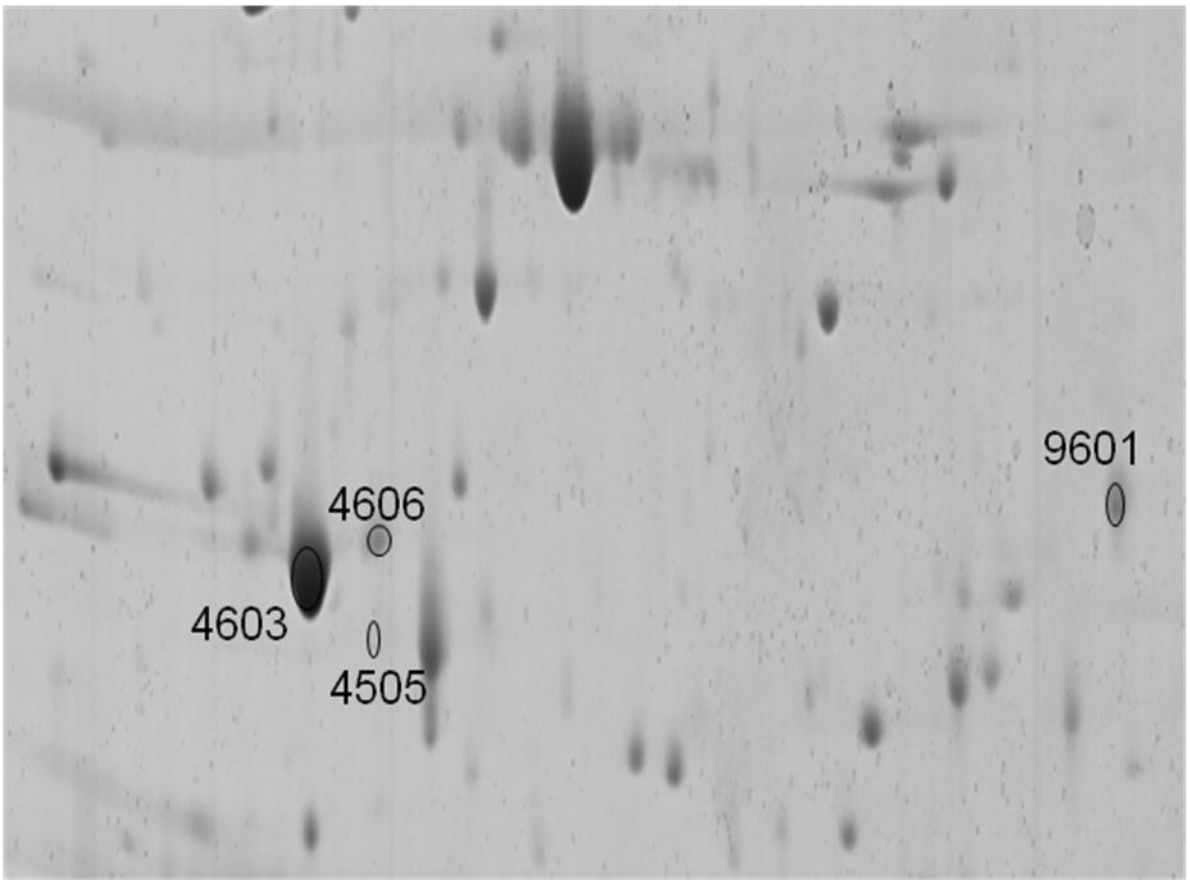


Figure 32: Figure showing the differentially expressed spots of outer membrane proteins in *E. cancerogenus.* SSP numbers 4606, 4505 and 9601 were identified to be outer membrane protein II and SSP numbers 4602 and 4603 were identified to be outer membrane protein A.

Outer membrane proteins are porins involved in transport of solutes, are immunogenic in nature and occur as epitopes on the bacterial surface exhibiting antigenic properties (Luo *et al.*, 1997). Outer membrane protein H have been known to be up-regulated during periods of iron starvation where lack of iron activates the *fur* gene which then activates the *ompH* gene leading to the increased production of the protein (Bosch *et al.*, 2001). These proteins may also be associated with intracellular trafficking, transport and secretion of vesicles apart from cell envelope biogenesis. They enable the adaptation of bacteria to various niches by obtaining iron from host-iron complexes. Increased amounts of glucose could also result in the suppression of the outer membrane protein H expression (Wheeler, 2009). Previous studies have shown that depletion of carbon sources in the growth media can result

in the activation of the gene encoding the synthesis of the protein since general expression in the presence of sugars is prevented by catabolite repression as studied in *Photobacterium* sp. (Bartlett & Welch, 1995). Outer membrane proteins A and II have been known to play an important role in conjugation, invasion of host tissues and detection of viruses apart from inducing immunogenic responses in the host (Singamsetty *et al.*, 2008). Deletion of the gene encoding the synthesis of the outer membrane proteins caused a reduction in the biomass, reduced survival rate at high temperatures, loss of viability in stationary phase (Beher *et al.*, 1980) and accumulation of low molecular weight solutes (Barrios *et al.*, 2006). Previous studies in *B. fragilis* (Ko *et al.*, 2009) and *Chlamydia pneumonia* (Hogan *et al.*, 2003) have shown the outer membrane potein encoding genes to be associated with virulence and seem to be up-regulated in severe cases of infection.

10.15 Intracellular trafficking, secretion and vesicular transport associated protein expression

Translocation associated TolB proteins were found to be differentially expressed in mucin enriched media in *E. cancerogenus* where up-regulation was observed in mucin Type III enriched media and down-regulation was observed in the mucin Type II enriched media. This protein was not identified in *B. fragilis*. TolB proteins are known to be involved in a number of functions that include translocation of colicins of group E and A including the delivery of the toxins to their respective target sites, assembly of porins in the outer membrane, recycling of peptidoglycan and linking the peptidoglycan layer to the outer membrane of the cell (Abergel *et al.*, 1999). They have been known to interact with trimeric outer membrane complexes including ompF and ompC but do not interact with monomeric forms like ompA (Rigal *et al.*, 1997). They are a part of the Tol dependent translocation system where they form complexes with porins and can initiate the import or export of colicin but increased expression of TolA proteins in the form of TolA II His-porin complexes can result in the down-regulation of the TolB proteins (Lazzaroni *et al*., 2002). They may also be essential for maintenance of the outer membrane stability (Lazzaroni *et al.*, 1999).

10.16 Expression of proteins involved in post translational modifications, protein turnover and chaperones

Putative thiol peroxidase was found to be differentially expressed in *B. fragilis* in mucin Type II and III enriched media but was not identified in *E. cancerogenus*. The protein was found to be up-regulated in SSP number 6301 and down-regulated in SSP numbers 4601 and 4605 in mucin enriched media in *B. fragilis* (Figure 33). . They play an important role as antioxidants and may be activated under conditions of oxidative stress and protect cells against the damaging effects of reactive oxygen species produced during metabolism (Missall *et al*., 2004). In the presence of hydrogen peroxide, they convert the reduced form of thioredoxin to the oxidised form and are also known as peroxiredoxins. The presence of varying concentrations of oxidants in the growth media produced as a result of the accumulation of toxic metabolic end products may be responsible for the differential expression of these enzymes in mucin enriched media. Under anaerobic growth conditions, the expression of these proteins is controlled by the binding of designated proteins that prevent the activation of the *tpx* promoter (Kim *et al.*, 1999). The regulation of these proteins may be attributed to the adaptation capabilities of bacteria to various ecological niches even under conditions of stress.

These proteins protect DNA and detoxify peroxides. The activation of the genes controlling the expression of these proteins is controlled by the *tpx* promoter (Kim *et al.*, 1999). Considering the lack of oxygen in anaerobic growth conditions, the down-regulation of these proteins suggests the possible lowering of oxidative stress in mucin enriched media (Kim *et al.*, 1999).

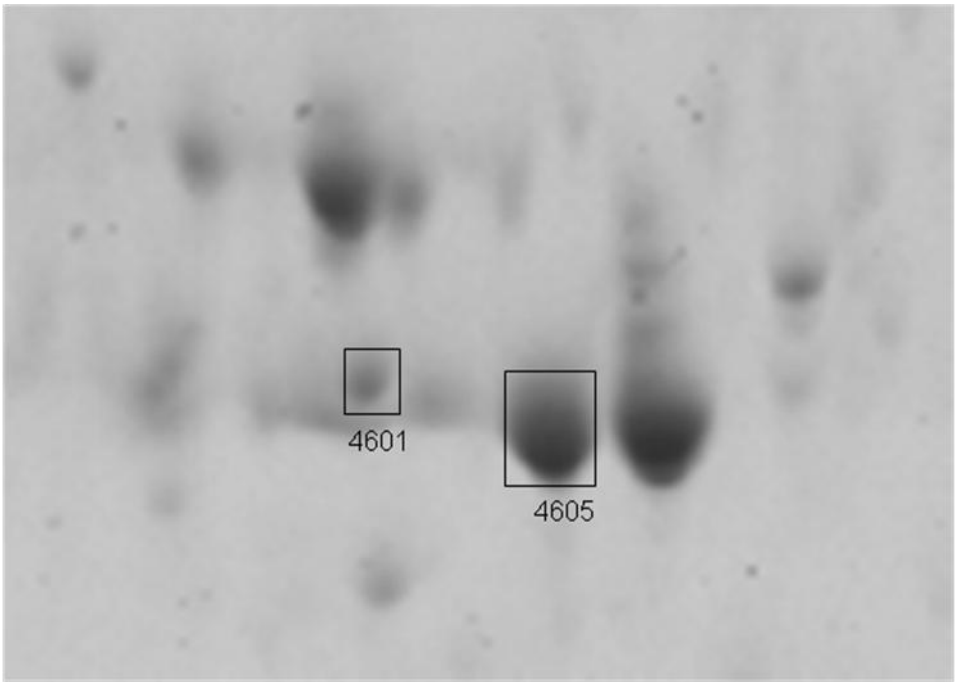


Figure 33: Figure showing the differentially expressed spots of putative thiol peroxidase in *B. fragilis*

E. cancerogenus showed a down-regulation of a protein from the glutaredoxin family in mucin Type III enriched media but was not identified in *B. fragilis*. Glutaredoxin related proteins are disulphide reductases that contain glutaredoxin and NADPH as cofactors. They have the presence of an intramolecular disulphide bond, act as electron carriers in the synthesis of deoxyribonucleotides and contribute to cellular functions like cell proliferation, viability and stabilisation during stress conditions (Holmgren *et al.*, 2005). Thiol disulphide oxidoreductases play an important role in disulphide bond formation in cytochrome biogenesis (Holmgren *et al.*, 2005). Previous studies in *E. coli* reveal the ability of these proteins to adhere to tissue culture cells and may contribute to the formation of pili (Donnenberg *et al.*, 1997). They have been known to contribute to bacterial virulence by acting as a folding catalyst suggesting that the down-regulation of these proteins in mucin media may be associated with the absence of a host immune response that induced expression (Yu & Kroll, 1999). This explains the reason behind the observation of a down-regulation of the glutaredoxin family related protein like thiol disulphide oxidoreductase in mucin Type III enriched media in *E. cancerogenus.* Other factors such as general metabolic requirements or toxic by product build up could also be responsible.

The co-chaperonin protein GroES was found to be up-regulated in mucin Type III enriched media in *B. fragilis*. This protein was not identified in *E. cancerogenus*. This is an oligomeric molecular chaperone which functions in protein folding and intracellular signals. It is a surface protein which is released from cells when required. It captures the substrate with the help of co-chaperone GroES and ATP which help with the binding. The substrate is then discharged into the microenvironment inside the chaperone promoting productive folding. The polypeptide is later released into the solution. GroES works as a co-chaperone with GroEL during protein folding. They have metal binding enzyme sites and can play an important role in host immunogenicity (Ranford *et al*., 2004). These chaperonins are known to act as intercellular signals interacting with a variety of cell types, including leukocytes, vascular endothelial cells and epithelial cells and may be responsible for key cellular activities such as synthesis of cytokines and adhesion proteins (Landry *et al*., 1996) . Co-chaperonin GroES is a known virulence factor and has been known to be associated with inducing inflammation and promoting cell proliferation (Lin *et al.*, 2006). Up-regulation of these virulence associated proteins in mucin Type III enriched media indicate that they may be expressed during infection of the host by bacteria.

A heat shock protein, GrpE was found to be differentially down-regulated in mucin enriched media in *B. fragilis* but no regulation was observed in *E. cancerogenus*. This is a stress related protein that works in conjunction with two other proteins namely DnaK and DnaJ forming complexes and is activated to repair damaged proteins or to initiate folding of proteins (Packschies *et al*., 1997). Apart from being involved in protein folding and disaggregation of heat shock proteins, these proteins also carry out the degradation of misfolded proteins. Unfolded proteins bind to DnaJ protein to form an unstable complex which is then stabilised by the hydrolysis action of DnaK. The movement of the folded protein is initiated by the removal of an ADP molecule which is converted to ATP coupled to the DnaK protein caused by the GrpE protein (Straus *et al*., 1990). Inactivation of the gene encoding the GrpE protein in *E. coli* has known to enhance the production of heat shock proteins and the continual production of the heat shock protein could not be stopped even though the growth conditions were returned to normal (Lipinska *et al.*, 1988). Down-regulation of

the protein indicates the lowering of heat shock or oxidative stress when grown in mucin enriched media when compared to the control media.

ATP synthase subunit E was up-regulated in mucin Type II and III enriched media in *B. fragilis* but no regulation was observed in *E. cancerogenus*. ATP synthases form proton channels that cause an influx of ions across the membrane thereby generating a proton gradient formed by the hydrolysis of ATP or oxidative phosphorylation of ADP and resulting in the generation of energy for metabolic processes (Feniouk & Junge, 2005). Differential expression studies in *Mycobacterium avium* showed the up-regulation of the the ATP synthase subunit C called the *AtpC* component of the ATP synthase complex (Radosevich *et al.*, 2007). The exact function of the subunit E is not known but the up-regulation of the protein may be associated with energy generation for efficient growth of the bacteria. Up-regulation of the ATPase subunit E may be associated with the availability of a large amount of sugar substrates for growth thereby supporting an increased generation of energy in *B. fragilis*. It is quite interesting that only the subunit E of the ATP synthase complex appeared to be up-regulated in mucin enriched media instead of the entire protein complex. ATP synthase subunit B, another component of the ATP synthase complex was identified during MS analysis of protein spots but did not show any differential expression in mucin enriched media. The exact reasons for the differential expression of the ATP synthase subunit E in mucin enriched media remain unknown.

10.17 Expression of proteins involved in energy production and conversion

Phosphoenol pyruvate carboxykinase (PEPCK), an important enzyme in the gluconeogenetic pathway catalysing the decarboxylation of oxaloacetate into phosphoenol pyruvate and carbon-di-oxide was shown to be down-regulated in mucin Type II and III enriched media in *B. fragilis*. This protein was not identified in *E. cancerogenus*. They utilise ATP or GTP as a source of phosphates and exhibit catalytic activity by binding to metal containing sites. The activation of these enzymes occurs when non fermentable sources of carbon are present in the media

and have been known to occur in ruminal and other anaerobic bacteria (Schocke & Weimer, 1997). Over expression of these proteins could affect the growth balance since they are known to have an influence over the oxygen concentrations in the media and this alteration in expression has been studied in *E. coli* (Chao & Liao, 1993). Since mucin acts as a rich source of carbon substrates, the fermentation of these substrates probably depend on the energy conservation and nutritional specialisation of *B. fragilis*. Studies have shown the up-regulation of the *acuF* gene which encodes the synthesis of PEPCK when non fermentable sources of carbon are present in the growth media (Hynes *et al.*, 2002). Hence the down-regulation of the protein in mucin enriched media can be attributed to the lack of requirement of more sugars for growth as mucin itself acts as a major source of carbon.

One of the interesting results from my study showed the differential regulation of malate dehydrogenase enzymes in *B. fragilis* but was not identified in E. *cancerogenus*. These enzymes were found to be down-regulated in mucin Type II and III enriched media in SSP number 4607 but the same enzyme was found to be up-regulated in mucin Type II and III enriched media in SSP number 4804 (Fig. 34). This could indicate that the enzyme exists in isoforms that may be differentially expressed based on their function but has been observed only in eukaryotes and is not known to exist in *B. fragilis*. Malate dehydrogenases catalyse the conversion of malate to oxaloacetate in the citric acid cycle and their expression has been known to vary with the carbon substrates and growth conditions used for culturing bacteria (Minarik *et al.*, 2002). Apart from being involved in carbohydrate metabolism, they may also be associated with amino acid transport and metabolism. Previous studies in malate dehydrogenase enzymes have shown that they exist as isoenzymes and also exhibit allelic polymorphisms in eukaryotes (Gietl, 1992) and in phototropic bacteria *Rhodobacter sphaeroides* and *Rhodopseudomonas palustris* (Eprintsev *et al*., 2008) . There were considerable amounts of distinctive amino acids associated with the allelic polymorphisms of the *mdh* gene indicating that this could have been caused by the horizontal transfer of genes in *E. coli* and *S. enterica* (Boyd *et al.*, 1994). The *mdh* gene in *E. coli* encodes for the ArcA protein whose expression is down-regulated under anaerobic conditions and regulation of expression was studied in response to varying oxygen, carbon and haeme availability in the growth media

(Park *et al*., 1995). The differential expression of these proteins may be associated with the presence of mucin in the semi-defined growth media which acts as a rich source of carbon providing a better survival rate for the bacteria.

Figure 34: Figure showing the differentially expressed malate dehydrogenase spots in *B. fragilis*

NAD dependent aldehyde dehydrogenase enzyme was found to be down-regulated in media containing mucin Type II in *E. cancerogenus*. These proteins were not identified in *B. fragilis*. The aldehyde dehydrogenase enzymes are very closely associated with the gamma glutamyl phosphate reductase enzymes. They catalyse the dehydrogenation or oxidation of aldehydes to produce acidic end products in the presence of NADP as a co-factor. One of the important features of this enzyme is that it can act as an allergen that induces IgE responses in humans and mammals. The enzyme may be acting as a potential virulence factor by triggering immune responses within the host by activating the IgG antibodies. Under anaerobic conditions, aldehyde dehydrogenases metabolise endogenous acetaldehyde and ferment nutrients to produce energy (Mikulskis *et al*., 1997). Previous studies showed that the expression of these proteins in germ free mice were modulated when challenged by bacteria and were found to be down-regulated (Fukushima *et al.*, 2003). Down-regulation of aldehyde dehydrogenase in mucin Type II enriched media indicates that the presence of mucin may be modulating the expression of this enzyme.

10.18 Expression of carbohydrate transport and metabolism associated proteins

Triose phosphate isomerase (TPI) which is involved in the interconversion of the TPI isomers dihydroxy acetone phosphate (DHAP) and glyceraldehyde-3-phosphate was found to be down-regulated in mucin Type II and III enriched media in *B. fragilis*. The enzyme plays an important role in carbohydrate metabolism through glycolysis and energy production. Regulation of the *tpi* gene which encodes the enzyme that can result in the accumulation of sugar moieties which becomes impossible to metabolise (Solem *et al*., 2008). They are also known to be immunogenic in nature where these enzymes were up-regulated in convalescent phase serum obtained from patients suffering from pneumococcal infections caused by *S. pneumoniae* (Zysk *et al.*, 2000). Growth of *B. fragilis* in cell culture could trigger an up-regulation of these proteins since they are associated with the host immune system. Differential expression studies in planktonic and biofilm cultures showed the up-regulation of these enzymes in biofilm conditions (Becker *et al.*, 2001). Previous studies have shown that these enzymes can be over expressed during stress conditions like deprivation of oxygen and this suggests that the down-regulation of these proteins may be due to the ability of bacteria to grow well in the presence of mucin (Yamaji *et al*., 2004). Regulation of expression of these proteins indicates that they play a vital role in carbohydrate metabolism and immune response suggesting that they could act as potential virulence factors when grown in host cells.

Up-regulation of the conserved domains of enolase or phosphopyruvate hydratase was observed in *E. cancerogenus* media enriched with mucin Type II and III. The enzyme forms an integral part of the glycolytic pathway catalysing the conversion of 2-phosphoglycerate to phosphoenol pyruvate. The expression of these enzymes has been found to be enhanced under anaerobic conditions and mutants lacking the presence of the enolase gene have been unable to utilise gluconeogenetic and glycolytic carbon sources (Lemaire & Wesolowski-Louvel, 2004). Enolase has also been known to play an important role in tissue invasion and pathogenesis in Gram-positive bacteria and may be over expressed in infection conditions (Liu & Shih,

2007). The up-regulation of these proteins in *E. cancerogenus* suggests the possibility that they may be indirectly involved in pathogenicity and may act as potential virulence factors and are therefore worthy of further study.

Phosphoglycerate kinase; a key enzyme involved in the glycolytic pathway was up-regulated in mucin Type II and III enriched media in *E. cancerogenus*. This protein was not identified in *B. fragilis*. These enzymes are involved in the anaerobic fermentation and generation of energy from carbon sources by catalysing the conversion of 1, 3-bisphosphoglycerate into 3-phosphoglycerate in the glycolytic cycle. Inactivation of the *pgk* gene encoding the synthesis of phosphoglycerate kinase resulted in the accumulation of 1, 3-bisphosphoglycerate thereby leading to the inability to generate energy for metabolism (Nakano *et al.*, 1999). Up-regulation of these enzymes indicates the efficient metabolisation of sugars which is probably enhanced in the presence of mucin which acts as a rich source of carbon.

Fructose 1,6-bisphosphate aldolase was found to be down-regulated in mucin Type II and III enriched media in *B. fragilis* following analysis using the PDQuest gel analysis software. This protein was not identified in *E. cancerogenus*. This enzyme catalyses the conversion of fructose 1,6 bisphosphate to dihydroxyacetone phosphate and glyceraldehyde-3-phosphate in the glycolytic pathway involving the binding of a divalent metal ion (Zinc) to the active site (Anon, 2009). This enzyme has been known to be expressed when B. fragilis cell suspensions were grown in media containing glucose and lactate as carbon sources (Macy *et al*., 1978). Previous studies in Trichomonas vaginalis revealed the up-regulation of malate dehydrogenases and fructose 1,6 bisphosphate aldolases and were found to be associated with virulence (Cuervo *et al*., 2008). Hence these enzymes may be potential virulence factors and play a vital role in the adherence of bacteria to host cells. They were found to be over expressed in the viable but non culturable state of bacteria suggesting that they play a vital role in the alternate metabolic pathways involved in the generation of energy (Heim *et al*., 2002) . In our experiments, fructose 1,6 bisphosphate aldolase was found to be down-regulated in mucin enriched media indicating better energy generation and viability of cells in mucin enriched media.

10.19 Amino acid transport and metabolism associated protein expression

Proteins from the carbamate kinase family involved in ATP and amino acid synthesis pathways was found to be up-regulated in mucin enriched media in *E. cancerogenus*. This protein was not identified in *B. fragilis*. Carbamate kinases catalyse the conversion of carbamoyl phosphate and ADP to carbamate and ATP. Hence this acts as a major source of ATP for energy metabolism. Previous studies have shown the over expression of this enzyme induced by growth of bacteria in anaerobic conditions, moderate expression induced by the presence of arginine and repression caused by the presence of ammonia (Abdelal *et al.,* 1982). Expression of carbamate kinase was found to increase when *S. aureus* cultures were grown under biofilm conditions and seem to play an important role in the growth of bacteria (Resch *et al.*, 2005). This indicates that the enzyme is involved in energy production reactions for the better growth of bacteria and its up-regulation may be attributed to the availability of rich sources of carbon under strict anaerobic conditions. Up-regulation of these proteins in mucin enriched media indicates an increase in energy metabolism in the presence of mucin in *E. cancerogenus.*

One of the other proteins that were differentially under expressed in mucin Type II and III enriched media was the lactoylglutathione lyase enzyme in *B. fragilis*. This protein was not identified in *E. cancerogenus*. This enzyme has been known to play an important role in the detoxification of the bacterial system and may be produced in response to oxidative stress (MacLean *et al.*, 1998). Only a general prediction of the function of these proteins is available associating them with amino acid transport and metabolism. This enzyme is also known as glyoxalase I and it catalyses the isomerisation of hemithioacetal adducts which are produced as a result of the reaction between glutathione (containing a glutathionyl group) and methylglyoxal (containing an aldehyde group). The end product of the reaction was lactoylglutathione which protected the system from the toxic effects of methylglyoxal (Korithoski *et al.*, 2007). The glutathione conjugates formed from the glyoxalase I-II pathway can activate KefB and KefC potassium channels that cause a lowering of the pH of the bacterial cell intracellularly thereby protecting them against the toxic effects of electrophiles (Ferguson *et al.*, 1998). Studies have shown that the genes encoding the expression of these enzymes when over expressed, resulted in an increased tolerance of the organism to methylglyoxal and osmotic stress exhibiting a

protective function (Takatsume *et al.*, 2005). They may be over expressed in cases of oxidative stress especially when acidic conditions prevail in the growth media (Wilkins *et al.*, 2002). Similar results have been observed in *S. mutans* where the *lgl* gene was over expressed during growth in acidic media (Korithoski *et al.*, 2007). But the down-regulation of the enzyme in mucin enriched media indicates a decrease in toxic metabolites or oxidative stress which suggests better viability and survival rates in mucin media rather than the control semi-defined media.

Proteins from the 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase I (DAHP) family was found to be up-regulated in mucin enriched media in *E. cancerogenus*. This protein was identified to be 2-dehydro-3-deoxy phosphooctonate aldolase and was not expressed in *B. fragilis*. 2-dehydro-3-deoxy phosphooctonate aldolase belongs to the transferase enzyme family and has been known to be up-regulated at transition stages from logarithmic to stationary phase in growth cultures (Sowell *et al*., 2008).The enzyme, 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase I plays an important role in the aromatic amino acid biosynthesis and shikimate pathway (Panina *et al.*, 2001). Two of the main amino acids synthesised using the shikimate pathway are tyrosine and phenylalanine and the feedback inhibition caused by the end products of the shikimate pathway does not affect the activity of DAHP synthase enzyme in *Bacillus sp*. (Kim, 2001). It catalyses the condensation reaction that converts phosphoenol pyruvate and D-erythrose-4-phosphate to 3-deoxy-D-arabino-heptulosonate 7-phosphate and this step plays an important role in the synthesis of amino acids like tyrosine and phenyl alanine (Herrmann, 1995). Since the substrates used in the condensation reaction are mainly sugar phosphates and mucin acts as a rich source of carbohydrates and complex polysaccharides, the up-regulation of these enzymes in *B. fragilis* can be understood.

10.20 Expression of poorly characterized proteins with a general function prediction only

The periplasmic protein, ecotin was found to be down-regulated in *E. cancerogenus* but not identified in *B. fragilis*. The protein acts as a broad range serine protease inhibitor that is capable of acting against enzymes like trypsin, chymotrypsin, blood peptidase factor Xa, thrombin and urokinase type plasminogen activator (Mcgrath *et al*., 1995). They exhibit a high degree of variation and develop adaptive interactions that suit the survival of the pathogen in host tissue and studies show that its location in the cell is periplasmic (Mcgrath *et al.*, 1991). They have a protective function and may be responsible for the bacteriostatic activity of the bacteria against neutrophil elastases (Eggers *et al.*, 2004). Since the presence of a host immune response triggers the expression of ecotin, its down-regulation suggests the lack of host digestive enzymes. This indicates that ecotin may be used as a protective system against proteolytic enzymes or other antibacterial agents and the lack of any components that challenge the growth of bacteria may have caused its down-regulation in mucin enriched media. In fact, media containing mucin support better growth of bacteria.

Differential expression of ribonucleotide reductase, pyruvate formate lyase (PFL) and autonomous glycyl radical cofactor were observed in *E. cancerogenus*. This protein was not identified in *B. fragilis*. Proteomic analysis of the spots revealed up-regulation of the protein in mucin Type II enriched media whereas down-regulation was observed in mucin Type III media. The autonomous glycyl radical cofactor plays an important role in the in the regulation of transcription and may be induced under conditions of oxidative stress since it acts as an antioxidant (Wyborn *et al.*, 2002). The enzyme ribonucleotide reductase is involved in the aerobic glycolysis reaction whereas pyruvate formate lyase takes part in the anaerobic glycolysis reaction (Leppanen *et al.*, 1999). Under oxidative stress conditions, the fragmentation of pyruvate formate lyase takes place and this induces the activation of the autonomous glycyl radical cofactor gene which then acts as an independent glycyl radical carrier and replaces the C terminal glycyl radical in PFL (Wagner *et al.*, 2001). Previous studies have shown that the enzyme pyruvate formate lyase is over expressed in the presence of pyruvate or milk and can yield mixed acid end products following anaerobic fermentation (Derzelle *et al*., 2005). Since the exact composition

of mucin remains unknown, the end products of glycolysis may be inducing the expression of this protein in mucin Type II enriched media when compared to the mucin Type III enriched media as revealed in previous work (Sawers & Bock, 1988).

A transcription elongation associated protein, elongation factor GreA was identified to be down-regulated in mucin Type II and III enriched media in *E. cancerogenus*. They are involved in the process of transcription regulation where the elongation factors cleave 2 or 3 nucleotides from the N terminal end of the nascent transcripts of RNA polymerases which are trapped by the arresting sites in the DNA forming locked DNA/RNA polymerase tertiary complexes as they pass through them. Once cleaved, resumption of the process of elongation takes place from the 3' end *(Lu et al*., 1997). Their functions include DNA dependent regulation of transcription, DNA binding and transcription elongation regulator activity (Stepanova *et al.*, 2007). These are cell surface associated proteins that are over expressed during stress conditions. Induced stress conditions like growth in lower pH (5.0) in *Streptococcus mutans* (Len *et al*., 2004) and mutation of the *pgi* gene that encodes the synthesis of the phosphoglycerate kinase enzyme trigger the increased expression of these proteins (Kabir & Shimizu, 2003). The lack of any stress conditions and the ability to survive effectively in the media could be responsible for the down-regulation of this protein in mucin enriched media when compared to the control non mucin media in *E. cancerogenus*.

10.21 Expression of signal transduction mechanism associated proteins

The universal stress proteins were found to be up-regulated in mucin Type II and mucin Type III enriched media in *E. cancerogenus*. This protein has not been identified in *B. fragilis*. This result is similar to the autonomous glycyl radical cofactor differential expression suggesting reduced accumulation of toxic end products or oxidative stress in mucin Type III media when compared to mucin Type II media. These cytoplasmic proteins play a major role in protecting DNA against damaging agents and they have been known to be over expressed when cells enter a preparatory phase before transition from the log phase to the stationary phase (Nystrom & Neidhardt, 1994). Expression may also be induced by depletion of carbon sources in growth media (Persson *et al.*, 2007) or in the presence of stress

inducing agents like salt or toxic by-products of metabolism (Weber & Jung, 2002). In some cases, these proteins have been known to be expressed even before complete depletion of carbon sources in the media (Chang *et al*., 2002). Previous studies in the universal stress proteins expressed by *E.coli* showed that 6 different proteins which were encoded by 6 different genes *uspA, B, C, D* and *E* were produced under different conditions of stress. The genes usp C and E were found to produce proteins that were involved in cell adhesion and motility and this could be related to the up-regulation of these proteins in mucin enriched media (Nachin *et al*., 2005).

10.22 Nucleotide transport and metabolism associated protein expression

Three other proteins found to be differentially expressed were cytidylate kinase in *E. cancerogenus*, phosphoglycerate kinase in *E. cancerogenus* and the 50S ribosomal protein L7/12 in *B. fragilis*. Cytidylate kinase proteins were found to be up-regulated in mucin Type III enriched media and down-regulated in mucin Type II enriched media in *E. cancerogenus*. These are cytoplasmic proteins that play a vital role in pyrimidine and nucleic acid metabolism by catalysing the transfer of a phosphate group from ATP thereby converting dCMP to dCDP and ADP. The two main functions of the protein include cytidylate kinase activity and ATP binding and they also require the binding of metal ions like magnesium (Mg^{2+}). When mutants lacking the conserved serine residues were generated, there was a dramatic decrease in the phosphorylation of dCMP indicating that these residues were highly conserved (Bertrand *et al.*, 2002). Not much information is available regarding the differential expression of these proteins but over expression in mucin Type III media suggests the possible availability of more energy in the form of ATP for nucleic acid synthesis unlike the mucin Type II or non mucin media.

E. cancerogenus showed the up-regulation of a periplasmic protein with a phospholipid binding domain in mucin enriched media. SSP number 6408 also showed a differential expression of a periplasmic protein disulphide isomerase I. This protein showed a down-regulation in the mucin Type II enriched media in *E. cancerogenus*. This protein was not identified in *B. fragilis*. The periplasmic protein disulphide isomerase belongs to the thiol disulphide oxidoreductase protein family

and is known to play an important role in the oxidative protein folding pathway. Gram-negative cell walls have the presence of phospholipids associated with their peptidoglycan and periplasmic layers. Periplasmic proteins form a part of bacterial transport systems with a high affinity and solute binding extracellular site (Ito *et al*., 1981). They may also be involved in the formation and aggregation of porins that transport nutrients across the membrane but the exact function of these proteins remain unknown (Lazar & Kolter, 1996). Previous studies have shown that increased concentrations of sucrose is capable of inducing the over expression of these genes and mucin being a rich source of carbohydrates like sucrose could be responsible for the up-regulation of these proteins (Costerto *et al*., 1974).

10.23 Summary

Some of the differentially expressed proteins identified in *E. cancerogenus* and *B. fragilis* were found to be hypothetical proteins that showed the presence of unknown domains of function. These include the SSP numbers 5003 (hypothetical protein 03947), 6306 (hypothetical protein 20877) and 8205 (Yfaz family protein) in *E. cancerogenus* and SSP numbers 1104 (BF2494), 2804 (BF2494), 5205 (BF1203) and 6206 (BF0301) in *B. fragilis*. Two of the proteins that were found to be differentially expressed in both *E. cancerogenus* and *B. fragilis* were elongation factor Ts and outer membrane proteins.

Apart from the differentially expressed proteins, several other proteins were found to be expressed in *B. fragilis* and *E. cancerogenus.* These can be classified as proteins required for the general metabolism and growth of bacteria in both control and mucin enriched media. A list of these proteins is available in the ‘Result section Table numbers 20 and 22 ’.

10.24 Future work

Cell culture coated multiwell plates can be used to study the interactions between the bacteria and eukaryotic cells. This may help in the better understanding of the virulence factors involved in the host infections caused by the bacteria since the presence of a host immune response may trigger the expression of virulence associated proteins.

The database used for searching and identifying the proteins of interest from the mass spectra was predominantly NCBI. It may be better to use other protein identification databases like PRIDE for determining the identity of proteins since they provide an identity based on searches in several databases including NCBI and UniProt.

A non gel based approach may also be used in studying differential expression in mucin media in both log and stationary phase. Even though the log phase is known to be associated more with virulence, it may be interesting to find out the differences in regulation that occurs in the stationary phase. One of the modern proteomics technologies includes Shotgun proteomics or Multidimensional protein identification technology (MUDPIT) where the proteins are subjected to a 2D-LC separation through strong cation exchange columns in the first dimension. The second dimension separation employs reverse phase chromatography. The cell lysates are tryptic digested before separation and pre-fractionation is carried out with iso-electric focussing. In order to obtain a better resolution of proteins, affinity chromatography using non-specific dyes could also be performed. The digestion and separation of proteins is followed by mass spectrometric analysis. The peptides are subjected to MS/MS by using electrospray ionisation (ESI) or matrix assisted laser desorption/ ionisation (MALDI). The spectrum generated from the mass spectrometric analysis is compared to a database to determine the identity of the protein of interest. This is considered an efficient method because it helps to identify proteins with high molecular masses, extreme pI, hydrophobic or when they occur in low abundance. This approach could also help in the identification of integral membrane proteins and overcome some of the drawbacks of two dimensional gel electrophoresis. Analysis of proteins using this technique could be used as a future method of studying differential expression of proteins.

In the proteomics technique, the sample processing step plays a very important role since slight variations can cause drastic changes in expression. In my methodology, the use of the clean up kit seems to help in obtaining a good resolution of proteins but the components of this kit remains unknown. It would be interesting to compare the expression of proteins when the acetone precipitation method is used instead of the 2D clean up kit.

Further work could also be done in trying to generate mutants that lack the genes associated with the expression of proteins like the outer membrane surface antigens which are known for their virulence and appear to be differentially expressed in both *B. fragilis* and *E. cancerogenus*. This could be useful in providing more information regarding the pathogenicity of these bacteria.

11. Results and discussion of the expression and structural analysis of *B. fragilis* proteins from glycoside hydrolase families 97 and 95.

11.1 Construction of plasmid vector expression constructs

NB. Expression vector constructs carrying the genes encoding proteins CAH09443 and CAH06598 from glycoside hydrolase families 97 and 95 respectively, were provided by Lee Ling Bong and Cheun Hong Yeap from Northumbria University.

FASTA format of the putative α-fucosidase (CAH06598) of glycoside hydrolase family 95 from *B. fragilis* NCTC 9343:

>gi|60491840|emb|CAH06598.1| conserved hypothetical protein [*Bacteroides fragilis* NCTC 9343]

```
  1 mkiklllllc cglwsscnsy dycpvtpses dlvftglars wdeamplgna tvgalvwqrd
 61 stlrlsldrt dlwdlrpvds lsgdnfrfsw vkehirqkny lpvqkkldwp ydmnpapski
121 pgaaiefple qigtptqvrl ylnnalcead wadgtqmqtf vhatepigwf vfrnlktpie
181 psiitpvynk tkpdgsldpv sgqdlhrlgy qqgkvvregn qityhqkgyg dfsydvtvcw
241 kqegetlygt wsvtsslsge qasekaeaal qrglkhdyqa hleywdkywa qssitlpdsv
301 lqkqyqnemy kfgsttrehs ypislqavwt adngklppwk gdyhhdlntq lsywpaytgn
361 hltegmgyln tlwnqrdayk rytrryfgte gmnipgvctl tgepmggwiq ysmsqtvaaw
421 laqhfylqwk ysadrtflke raypfikdva iyleqisevt pegvrklefs sspeifdnsl
481 qawfsdmtny dlammhflfk atselaheln ladeaghwas leaqlpdydi deegcltfak
541 gypykeshrh fshamaihpl glidwsdgek sqhiiratlk rldkvgpdyw tgysyswlan
601 mkarafdgeg aaqalktfae cfclkntfha ngdqtqsgks rftyrpftle gnfafaagiq
661 emllqshtgv irifpaipke wkdvsfenlr amgaflvsar meggeinrvr iysekggmlk
721 marpgtlkpn knytlsgtdi lnidtqagew ielnp
```

The protein was found to contain a signal peptide that was cleaved between the 19th and 20th position of the amino acid sequence (Appendix I). The protein parameters were determined from the Protparam tool (Refer to Appendix I for details).

FASTA format of the putative α-glucosidase (CAH09443) of the glycoside hydrolase family 97 from *B. fragilis* NCTC 9343:

>gi|60494642|emb|CAH09443.1| putative exported protein [*Bacteroides fragilis* NCTC 9343]

```
  1 mkrkmmslll alavisgssv yakvidvmsp ngaikvsvdi kdriyysvsy dndqllkdcy
 61 lnlqlqnetl gtnphlrstk rgtidesvkr eipfknaivr nhcntlrmnf sgnyavefrv
121 fdngiayrfv tdkkgdnivm gedfainfpt nykahlsqpd gfktsyecpy thvdtekyaa
181 tdrmsylpvl ietdkaykil iseadlsdyp cmflkstgkn gmqsifpkap lafgedgdrs
241 lkiteeadyi aktdgkrsfp wrmmvisked kelienemvy nlsapcvled yswikpgqvs
301 wewwhdarly gvdfrsgfnm dsykyyidfa skfgipyiim degwakntrd pftpnptinl
361 telikygkdr nvkivlwlpw ltvenhfdlf ktfadwgiag vkidfmdrsd qwmvnyyerv
421 akeaakhklf vdfhgafkpa glerkypnvl syegvlgmeq ggnckpensi ylpfmrnavg
481 pmdftpgsmi saqpednrst ranamgsgtr afqmalfiif esglqmladn pvyyyrelpc
541 tefitsvpvt wdetkvlyak vgeavvvakr kgeqwfiggi tgnqpqniei dlgfipagqs
601 ftltsfedgi nadrqamdyk kkestvnnqt rmtlkmvrng gwagtikmk
```

The SignalP server predicted the presence of a signal peptide with probable cleavage positions between the 22nd and the 23rd residues of the amino acid sequence (Appendix I). The protein parameters were determined from the Protparam tool (Appendix I).

The genes were amplified using the following primers:

Forward primer for BF3763

5'- ATGAAAGTGATTGACGTAATGTCTCCCAACGG-3'

Reverse primer for BF3763

5'- ATTGTTCCGGCCCATCCCCC-3' and

Forward primer for BF0855

5'- TATGACTATTGTCCGGTCACTCCTTCAGAGAG-3'

Reverse primer for BF0855

5’- AGGGTTCAACTCGATCCATTCACCGGCTTGAGT-3’

The amplified genes which encoded the mature CAH09443 and CAH06598 proteins were cloned into the pETYSBLIC vector (see Appendix C)

The BF3763 and BF0855 genes were cloned and expressed in *E. coli* using the pET-YSBLIC vector system. *E. coli* was used as a host for the expression of recombinant proteins because of the ease of promoter control, high amount of product yields, high densities of growth in cultures, easy to grow with regards to metabolic carbon requirements (Brown, 1953). The pET-YSBLIC vector is a modification of the pET28a vector developed at York University where the gene of interest can be inserted into the vector through ligation independent cloning. The LIC vector is modified to encode a hexa-histidine tag that is incorporated at the *N* terminus of the protein of the encoded gene thereby making its purification easy. These factors make the LIC vector the most suitable candidate for cloning and expression experiments.

The SignalP and TMHMM websites were used to determine the presence of signal peptides and transmembrane helices in the proteins of interest so that genes sequences encoding these protein elements could be excluded during cloning. This was done to prevent the accumulation of secreted and transmembrane proteins in the inner membrane of the host which would inhibit growth.

11.2 Protein expression comparison using IPTG induced media and auto-induction media

Figs. 35 and 36 show SDS gels of CAH06598 and CAH09443 proteins, respectively, isolated from recombinant *E.coli* cells grown in IPTG induced media and auto-induction media.

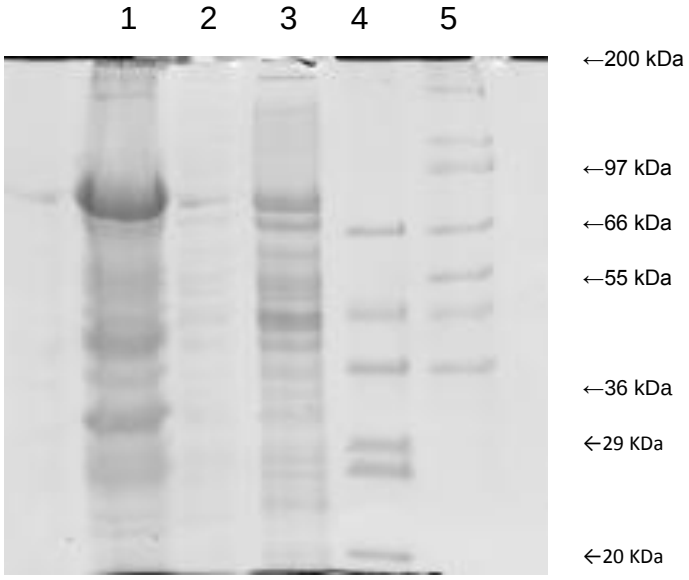


Figure 35: Analysis of CAH06598 expression (86 kDa) using a 12% (w/v) SDS gel obtained from growing the recombinant *E. coli* in IPTG induced media. Lane 1- Solubilisation buffer sample, Lane 2- 1 in 10 dilution of CFE, Lane 3- Cell free extract (CFE), Lane 4- Low molecular weight marker, Lane 5- High molecular weight marker.

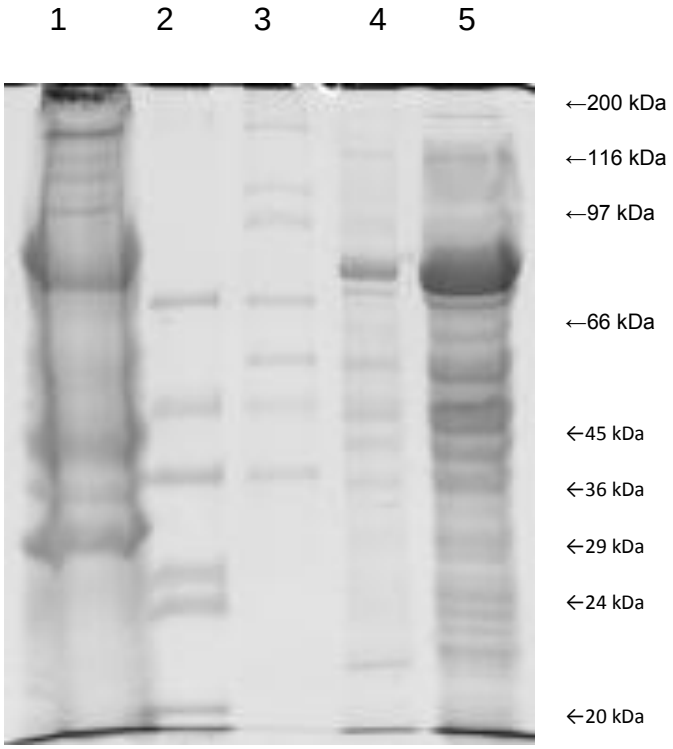


Figure 36: Analysis of CAH09443 expression (74 kDa) using a 12% (w/v) SDS gel obtained from growing recombinant *E. coli* in auto-induction media. Lane 1- Solubilisation buffer sample, 2- Low molecular weight marker, 3- High molecular weight marker, 4- 1in 10 dilution of CFE, 5- CFE

Since the cell free extract in the auto-induction media showed better expression than the IPTG induced media, the former was used in the protein purification experiment.

The analysis of protein production was performed by comparing the expression levels in auto-induction and IPTG induced media. The auto-induction media which had the presence of glucose and α-lactose in it supported the growth of the recombinant *E. coli* and glucose was initially used as the source of carbon. Once the glucose present in the media was completely depleted, the cells started metabolising the lactose present. According to previous studies, this stage was normally achieved when the cells entered their mid or late log phase (Studier, 2005). The cultures were incubated overnight at 30°C with constant aeration and harvested at their stationary phase of growth (OD >4.0). However, the differences in level of expression of the protein of interest was not studied in lower temperatures or at various time points post induction even though lower temperatures (20°C) have been known to support

better growth due to increased solubility of oxygen (Donovan *et al*., 1996). But since the expression of proteins was found to be very significant in the culture conditions used in the experiment and no other contaminating proteins were detected, the same system was used to obtain stable protein expression.

11.3 Protein expression and purification of *B. fragilis* proteins CAH06598 and CAH09443

CAH06598 and CAH09443 were extracted, purified using immobilised nickel affinity chromatography (Figs 37 and 41, respectively) and run on 12% (w/v) SDS-gels (Figs. 38 and 42, respectively) to confirm the proteins of the right molecular weight. Further purification of the proteins was performed using gel filtration and fractions observed on SDS gels.

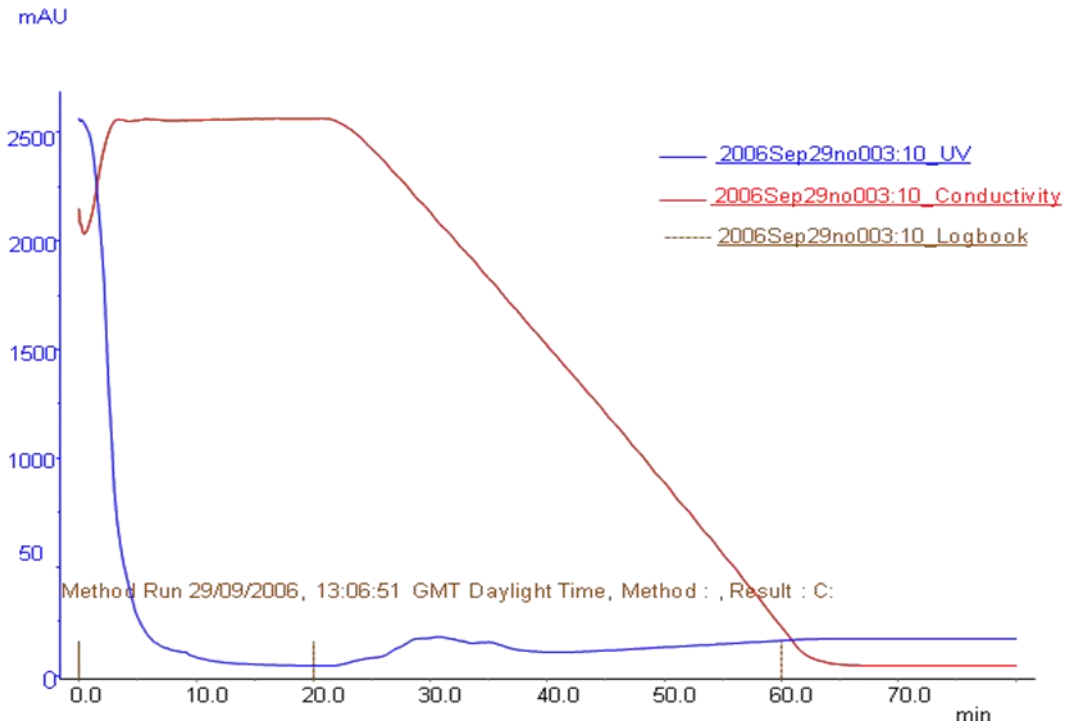


Figure 37: Nickel column purification of CAH06598 protein. The blue line represents the UV absorbance and the peak that appears between 20 and 40 min confirms the results from the SDS gels.

Figures 38: Nickel column purification results for CAH06598 protein

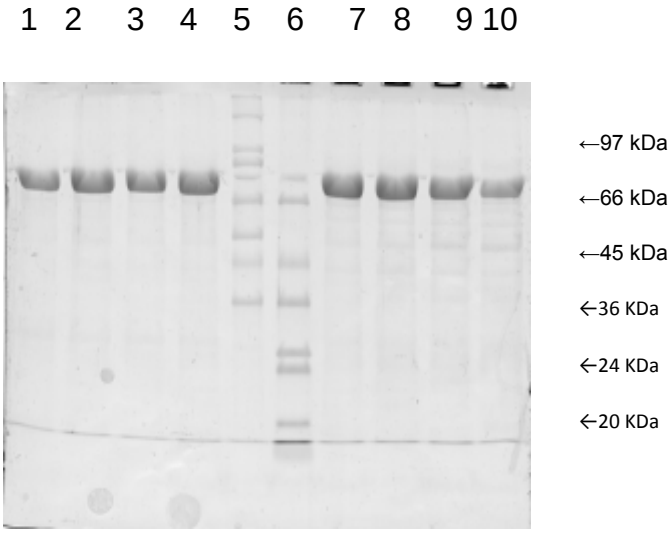
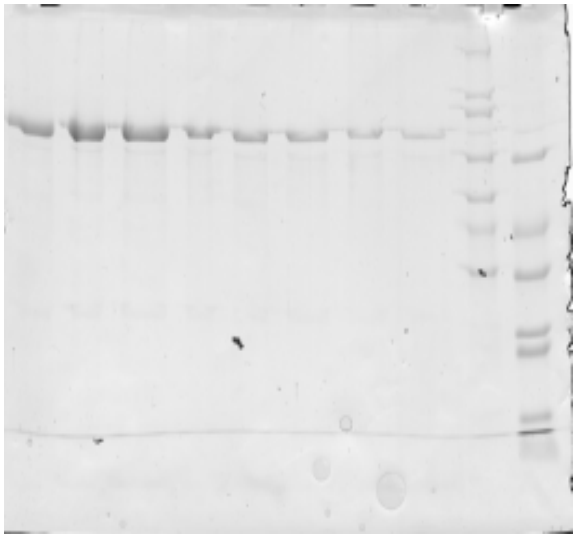


Figure 38.1: 12% (w/v) SDS gel showing purified fractions. Lanes 1, 2, 3, 4, 7, 8, 9 and 10 shows the presence of a band at 86 kDa indicating the presence of the CAH06598 protein which corresponded to the eluted fraction numbers of 26, 27, 28, 29, 30, 31, 32 and 33 respectively. Lanes 5 and 6 represent the high and low molecular weight standards respectively.

1 2 3 4 5 6 7 8 9 10



←200 kDa

←97 kDa

←66 kDa

←55 kDa

←45 KDa

←36 KDa

Figure 38.2: 12% (w/v) SDS gel showing purified fractions. Lanes 1 to 8 shows the decrease in intensity of the band at 86 kDa, indicating the presence of the CAH06598 protein, as elution continued and represent fraction numbers 34 to 41 respectively. Lanes 9 and 10 show the high and low molecular weight standards.

Fractions 34-41 were pooled and concentrated in a 30 kDa cut-off centrifugal concentrator and washed into 20 mM HEPES buffer pH 7.4. The concentrated protein was then purified using a HiLoad 16/60 Superdex 200 prep grade gel filtration column (Fig. 39) and the fractions containing CAH09443 analysed via SDS-PAGE (Fig. 40).

Figure 39: Gel filtration graph of CAH06598 protein showing the UV absorbance peak. Refer to Appendix section I for details.

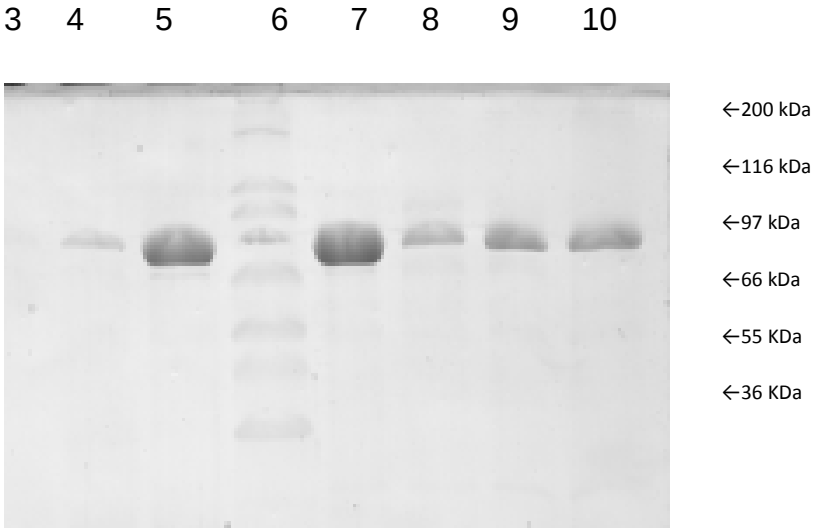


Figure 40: 12% (w/v) SDS gel showing the gel filtration fractions containing CAH06598 protein. Lanes 3, 4, 5, 7, 8, 9 and 10 represent the fractions 14, 15, 16, 17, 18, 19 and 20 respectively indicating the presence of protein bands at 86 kDa. Lane 6 shows the high molecular weight standard.

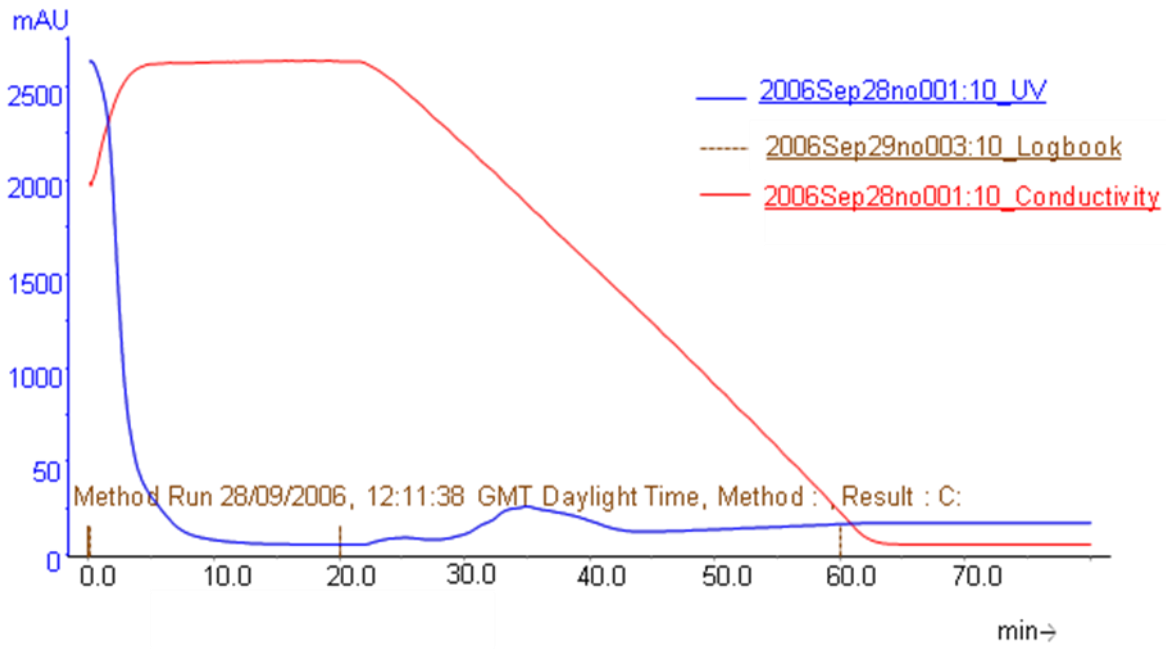


Figure 41: Nickel column purification results for CAH09443 protein. The blue line represents the UV absorbance and the peak that appears between 30 and 40 min indicates the elution of the protein at these corresponding fractions.

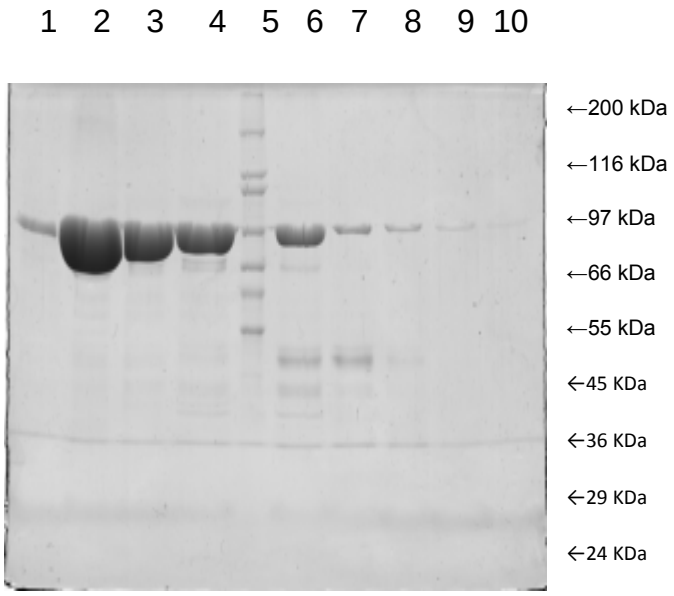


Figure 42: Nickel column purification results for CAH09443 protein (12% (w/v) SDS-gel). Lanes 2, 3, 4 and 6 represent fractions 34, 36, 38 and 40 respectively showing the presence of bands at 74 kDa. Lanes 1, 7, 8, 9 and 10 represent the fractions 31, 41, 43, 45 and 47 and Lane 5 represents the high molecular weight standard.

Fractions 34-40 were pooled and concentrated in a 30 kDa cut-off centrifugal concentrator and washed into 20 mM HEPES buffer pH 7.4. The concentrated protein was then purified using a HiLoad 16/60 Superdex 200 prep grade gel filtration column (Fig. 43) and the fractions containing CAH09443 analysed via SDS-PAGE (Fig. 44).

Figure 43: Gel filtration graph of CAH09443 protein showing the UV absorbance peak. Refer to Appendix section I for details.

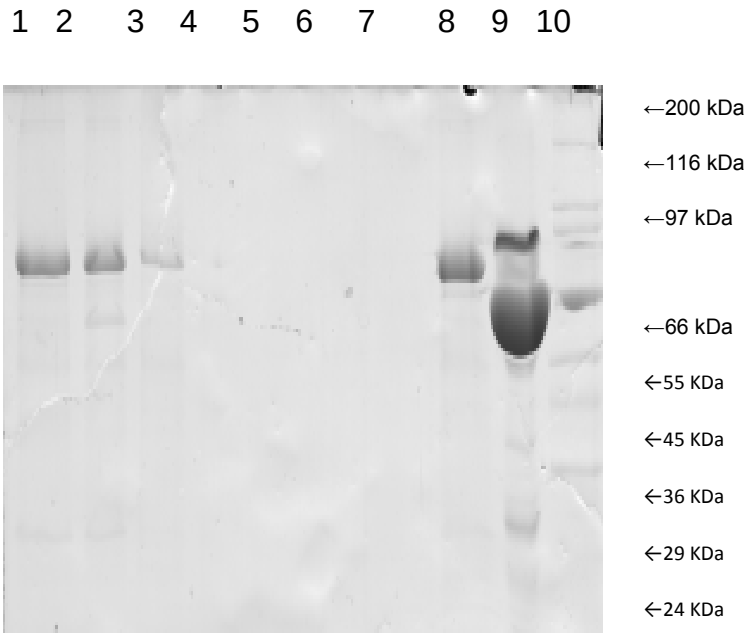


Figure 44: 12% (w/v) SDS gel showing the gel filtration results for CAH09443 protein. Lanes 1, 2, 3 and 8 show the fractions 10, 12, 13 and 15 respectively and Lane 9 and 10 represent the low and high molecular weight standards. Lanes 4 to 7 represent fractions 17, 19, 21 and 23 which do not show the presence of bands. Fractions 10 to 15 were pooled for further concentration.

The purified proteins were concentrated to a volume of 0.5 mL and the concentration of protein was estimated using Bradford's assay. They were diluted down to a concentration of 15 mg/mL from 43 mg/mL (CAH06598) and 29 mg/mL (CAH09443) respectively before crystallogenesis.

The proteins expressed and purified from CAH09443 showed tendencies towards precipitation at pH 9 and above. Hence the proteins were maintained in 5 mM HEPES buffer at a pH of 7.4. The nickel column purification of recombinant proteins from *E. coli* did not yield good peaks of absorbance at 280 nm in the UV chromatograms but strong and well detectable bands were observed on performing the SDS-PAGE analysis. The UV chromatogram peaks for GH95 proteins were found to be better resolved when compared to CAH09443.

.

11.4 Protein crystallogenesis and diffraction results for *B. fragilis* CAH06598 and CAH09443.

A number of different crystallisation screens were used in the attempt to crystallise *B. fragilis* CAH06598 and CAH09443 proteins. These include PEG/Ion, Hampton Screen (HS) I and II, Clear Strategy Screen (CSS) I and II, PEG/Anion, PEG/Cation, PEG/pH, SaltRx, Newcastle and Index screens.

Crystals of CAH06598 proteins were observed in the Peg/Ion screen 3 containing 0.2 M ammonium fluoride and 20% (w/v) polyethylene glycol (PEG) 3350, CSS II screen condition 35 containing 10% PEG 8000, 10% PEG 1000 and 0.2 M Calcium acetate, HS I screen condition 20 containing 0.2 M ammonium sulphate, 0.1 M sodium acetate pH 4.6 and 25% (w/v) PEG 4000, Newcastle screen condition 1 containing 50% PEG 400, 0.2 M lithium sulphate and 0.1 M sodium acetate pH 5.1 and PEG/Anion screen condition 8 containing 0.2 M sodium sulphate and 20% PEG 3350. But no diffraction was obtained from the crystals of the above mentioned conditions.

The PACT ANION screen number 12 for CAH09443 produced crystals that diffracted and the screen was composed of 20% (v/v) PEG 3350 (Polyethylene Glycol) and 200 mM of sodium malonate. The proteins were buffered in 5 mM HEPES at a pH of 7.4. Other conditions that produced crystals include Peg/Ion condition 40 containing 0.2 M potassium thiocyanate, 0.1 M Bis-tris propane pH 8.5 and 20% PEG 3350, PEG/Anion condition 34 containing 0.2 M sodium potassium phosphate, 0.1 M Bis-tris propane pH 7.5 and 20% PEG 3350 and PEG/Anion condition 22 containing 0.2 M sodium potassium phosphate, 0.1 M Bis-tris propane pH 6.5 and 20% PEG 3350. But unfortunately these crystals did not diffract.

Crystallisation studies were performed on both the CAH06598 and CAH09443 protein families but no crystals were detected in CAH06598 that gave a positive diffraction pattern. CAH09443, yielded crystals with a positive diffraction pattern in the PACT/ANION screen number 12 but when the selenomethionine proteins were produced in minimal media, crystals were obtained but did not show a diffraction pattern. Attempts were made at growing selenomethionine protein crystals by seeding existing crystals into new screen conditions but no detectable crystals were observed.

11.5 Production, purification and crystallogenesis of a selenomethionine derivative for *B. fragilis* CAH09443

A selenomethionine derivative of CAH09443 was expressed, purified (Fig. 45) and crystallised but no positive diffraction patterns were obtained from the crystals.

Figure 45: Purification results for selenomethionine incorporated CAH09443 derivative.

12% (w/v) SDS gels showing the nickel column purification and gel filtration purification results for the selenomethionine preparations of CAH09443 respectively.

1 2 3 4 5 6 7 8 9 10

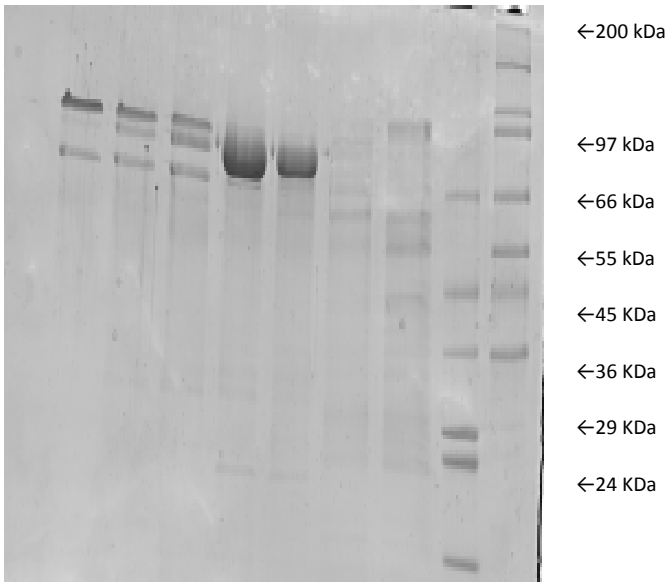


Figure 45.1: 12% (w/v) SDS gel showing the nickel column purification results for selenomethionine preparation of CAH09443 protein. Lanes 5 and 6 shows the presence of bands at 74 kDa and these were from fractions 41 and 43 and Lanes 1, 2, 3, 4, 7 and 8 represent fractions 31, 33, 36, 38, 46 respectively. Lanes 9 and 10 represent low and high molecular weight standard.

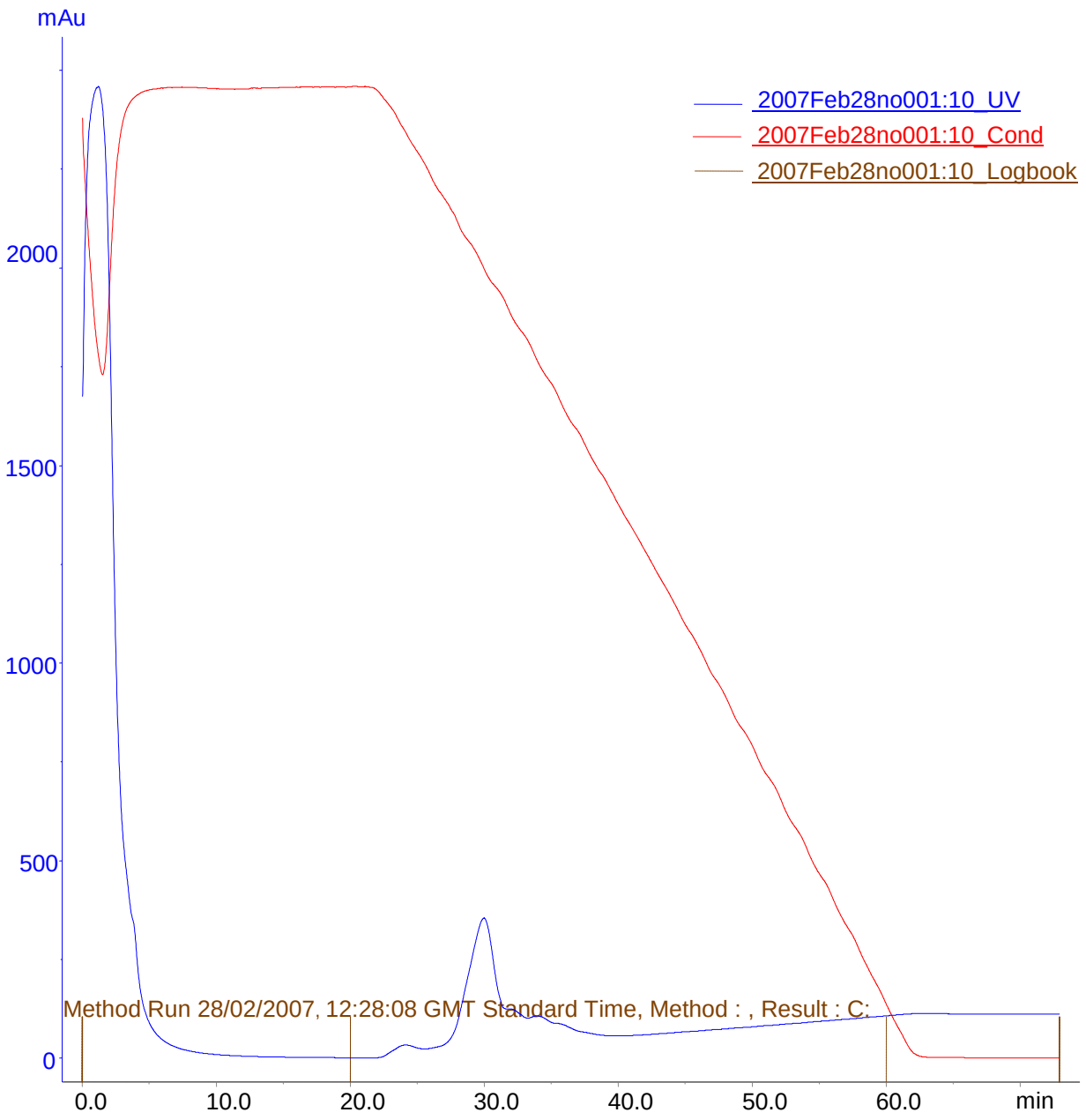


Figure 45.2: Nickel column purification results for the selenomethionine derivative of CAH09443 protein.

Figure 45.3: Gel filtration graph of CA09443 protein showing the UV absorbance peak for the selenomethionine derivative. Refer to Appendix section I for details.

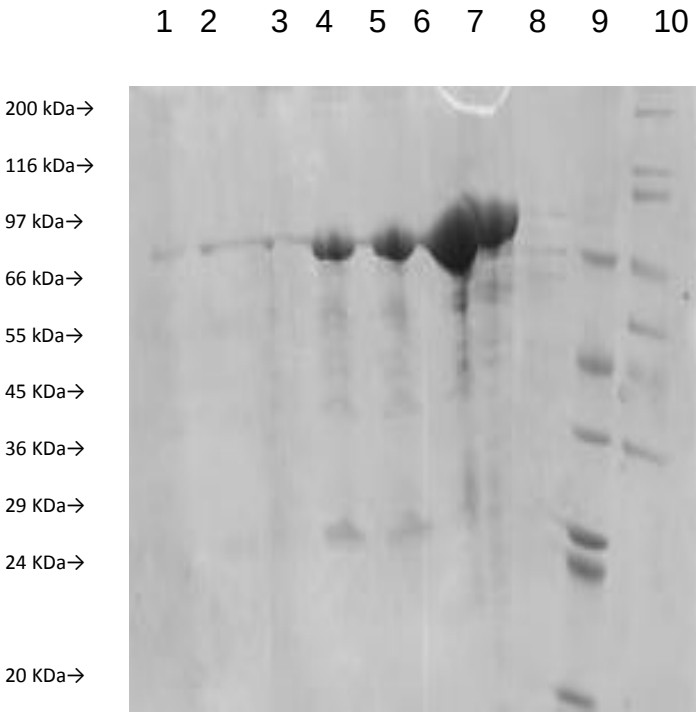


Figure 45.4: 12% (w/v) SDS gel showing gel filtration purification results for selenomethionine preparation of CAH09443 protein.

Lanes 4 to 7 shows the presence of purified fractions and the fraction numbers were 12, 13, 14 and 15. Lanes 9 and 10 were the low and high molecular weight standards and lanes 1 to 3 and 8 represent fractions 9, 10, 11 and 16 respectively.

The gels indicate the presence of bands at the correct molecular weight of 74 kDa.

Crystals were produced on screening using the Newcastle crystal screen condition 17 containing 40% MPD, 5% PEG 8000 and 0.1 M sodium cacodylate pH 7.0 but did not show diffraction.

11.6 Structure solution of CAH09443 via molecular replacement

NB. Structure solution of CAH09443 was performed by Dr Edward Taylor at York Structural Biology Laboratory.

The crystals were found to belong to space group P 2 21 21 with two molecule in the asymmetric unit with the approximate cell dimensions of a = 60.33 Å b = 132.82 Å c = 163.65 Å (Table 24). The structure was determined via molecular replacement using *Bacteroides thetaiotaomicron Bt*GH97b (UniProtKB/TrEMBL entry Q8A6L0) structure (PDB code 3A24) as a search model and refined to a resolution of 2.70 Å with an R-factor of 0.225 and R-free of 0.296 (Table 24).

Data Processing	***Bf*GH97**
Space Group (No.)	P 2 21 21 (18)
Unit Cell lengths (Å)	a = 60.33 b = 132.82 c = 163.65
Unit Cell angles (º)	α = β = γ= 90
Molecules in asymmetric unit	2
Resolution Range (outer shell)[1]	81.82- 2.70 (2.85-2.70)
R_{merge}*	0.11 (0.464)
< I/□$_I$ >*	6.3 (15.6)
Completeness*	99.8 (99.7)
Redundancy*	6.6 (5.7)
Refinement Statistics	
Resolution Range (Å)	103.13- 2.70
R_{cryst}	0.225
R_{free}[2]	0.296
No. protein atoms	9953
Mean ***B*** value protein atoms (Å^2)	28.0
Mean ***B*** value solvant atoms (Å^2)	29.4
Ramachandran Statistics[3]	**88.2% preferred regions, 7.5% allowed regions, 4.2% outliers**

Table 24: Data collection and refinement statistics for *Bf*GH97 (*B. fragilis* CAH09443 protein). Structure figures were drawn with PyMOL (DeLano Scientific www.pymol.org).

1 Numbers in parenthesis correspond to the high resolution outer shell

2 Estimated Standard Uncertainty, based upon R_{free}, calculated using REFMAC

3 Calculated using Validation options in COOT

11.7 Three dimensional structure of CAH09443.

The tertiary structure of CAH09443 reveals two domains (Fig. 46); an N-terminal β-super-sandwich domain (in yellow) and a canonical $(\beta/\alpha)_8$ barrel (in red). This is a similar architecture to other GH97 enzymes, specifically the *Bacteroides thetaiotaomicron* α-glucosidase, *Bt*GH97a (Gloster *et al*., 2008; Kitamura *et al*., 2008) and α-galactosidase, *Bt*GH97b (Okuyama *et al*., 2009). Previous studies in the *Bacteroides thetaiotomicron* GH97 showed that one of the general characteristics of GH97 enzymes may be their ability to possess two catalytic residues that act as bases at the end of β strands 3 and 5 or a conserved nucleophilic residue at the end of the β strand 4 (Refer to Fig. 47 for calcium binding sites of the *Bt*GH97 enzymes) (Okuyama *et al*., 2009). Our enzyme, BfGH97 seems to show the presence of a conserved nucleophilic aspartic acid residue at the position 405 which corresponds to the position 415 in *Bt*GH97b (Refer to Fig. 48 for details of the ClustalW alignment).

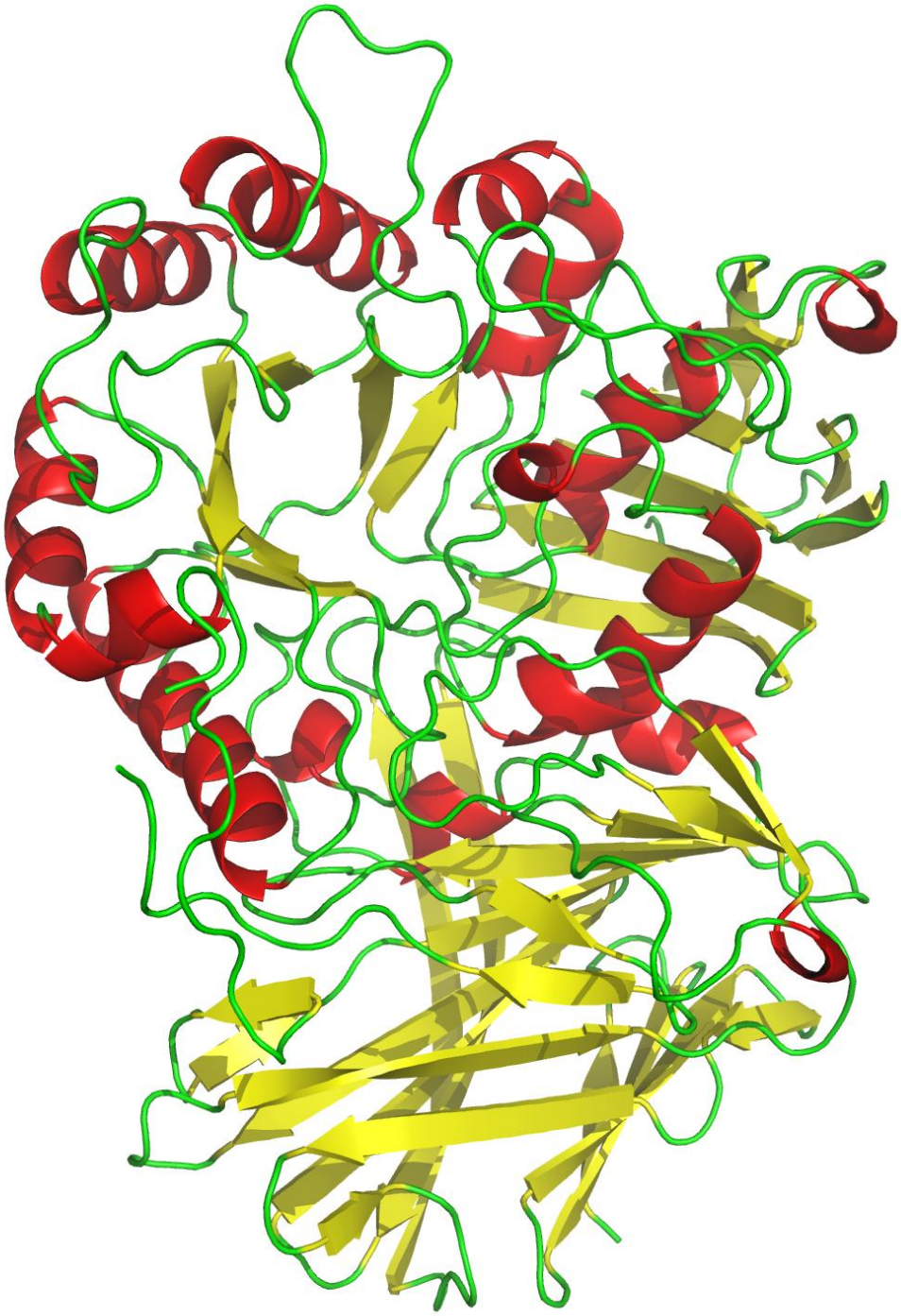


Figure 46: Structure of CAH09443. Crystal structure of CAH09443 coloured according to domains. The N-terminal β-super-sandwich domain is in yellow and a canonical $(\beta/\alpha)_8$ barrel is in red.

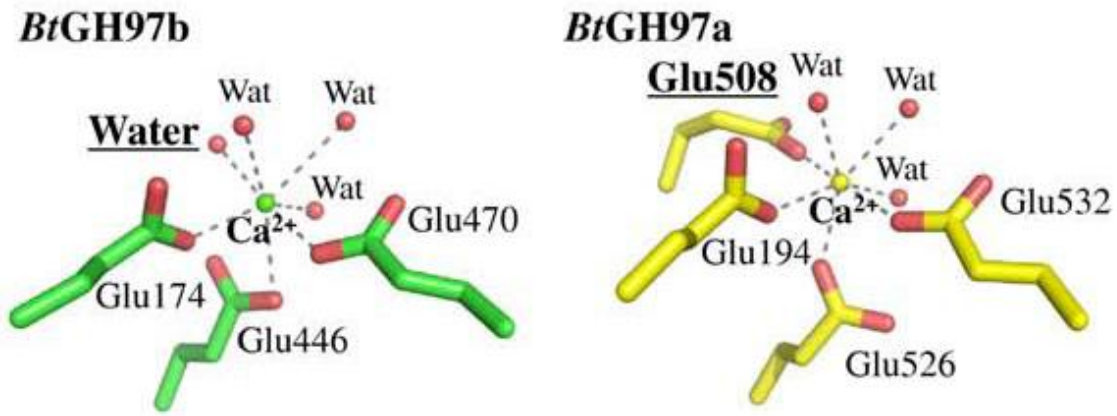


Figure 47: Figure shows the calcium binding sites in *Bt*GH97a and *Bt*GH97b (Okuyama *et al*., 2009).

11.8 Biochemical assay for CAH06598 and CAH09443.

CAH06598 was predicted to be a putative uncharacterized protein and CAH09443 was predicted to be α-glucosidase.

Table 25 shows the presence or absence of activity in fluorescent and chromogenic assays for CAH06598 and CAH09443 proteins against methylumbelliferyl and p-nitrophenyl glycoside substrates, respectively.

Substrates used in enzyme assay	***B. fragilis* CAH06598 (GH95)**	***B. fragilis* CAH09443 (GH97)**
Para nitrophenyl α-D-glucoside	-	+ (18 h incubation)
Para nitrophenyl α-D-fucoside	+ (18 h incubation)	-
Para nitrophenyl α-D-galactoside	-	+ (18 h incubation)
Para nitrophenyl β-D-maltoside	-	-
4-methylumbelliferyl α-D-glucoside	-	+ (2 min incubation)
4-methylumbelliferyl α-D-fucoside	+ (2 min incubation)	+ (2 min incubation)
4-methylumbelliferyl β-D-maltoside	-	-

Table 25: Determination of the activity of CAH06598 and CAH09443 against a variety of para nitrophenyl and methylumbelliferyl substrates at 37°C.

+ indicates presence of activity

- indicates absence of activity

Two types of enzyme assays were performed to determine the putative activity of the enzymes from CAH06598 and CAH09443 which was α-fucosidase and α-glucosidase respectively. The fluorimetric assays showed activity towards their respective substrates but the chromophoric assays did not show positive activity. As mentioned in the results section Table number 23, the pNP assay showed a change in colour of the reaction mixture when incubated overnight at 37°C but no instant reactions or change of colour to yellow was observed. This suggests that the reaction occurred at a slower rate in the pNP assay which could probably be optimised by changing the concentration of the substrates, enzymes or by modifying the reaction conditions. The enzymes were found to be very specific in their reactions towards the substrates even though a wide range of substrates were used. This also indicates that the fluorescent assays were more sensitive when compared to the chromophoric assays. The CAH06598 putative α-fucosidase enzyme exhibited activity towards 4-methylumbelliferyl α-D-fucoside and CAH09443 putative α-glucosidase exhibited activity towards 4-methylumbelliferyl α-D-glucoside indicating the specificity of the enzyme in its activity against the substrates used. No assay experiments were performed to determine the optimum temperature and pH for the activity of the enzymes since positive activity was observed at 37°C and the pH was maintained at 7.4 since the CAH09443 enzyme precipitated at higher pH.

α-Glucosidase enzymes are also known as acid maltases, glucoinvertases, glucosidosucrases, lysosomal α-glucosidases, maltases and maltase-glucoamylases. They are capable of hydrolysing terminal, non reducing 1→4 linked D-glucose residues resulting in the release of α-D-glucose as an end product. They are also known as exo-enzymes because they carry out exohydrolysis of α1→4 glucosidic linkages and hydrolyse oligosaccharides rapidly. The hydrolysis of polysaccharides is comparatively slower and industrial enzymes are capable of hydrolysing 1, 6 α-D-glucose linkages in polysaccharides (Hughes *et al.*, 2003).

These enzymes play an important role in the hydrolysis of complex polysaccharides into glucose and act as one of the major factors that contribute to type 2 diabetes. Attempts are being made to use acarbose, which is a competitive inhibitor of α-glucosidase to lower the blood glucose levels in diabetic patients following the consumption of food containing carbohydrates (Van de Laar *et al.*, 2005). The

enzymes also take part in glycogen metabolism and nutrient uptake of bacteria. They catalyse the transglycosylation of α-D-glucose moieties and are industrially useful in the biosynthesis of oligosaccharides or glycoconjugates (Shirai *et al.*, 2008).

11.9 Future work

Further experiments could be performed to generate a catalytic mutant of CAH09443 that lacks the putative acid catalyst, Asp404. This amino acid residue has been identified to be the acid catalyst as it aligns with the acid catalyst, Asp415 of the α-galactosidase, *Bt*GH97b, from *Bacteroides thetaiotaomicron* (Fig. 47). Once the mutant is generated, its structure can be solved in the presence of substrate, therefore providing additional structural data with respect to enzyme-substrate interactions.

```
gi|60494642|emb|CAH09443.1|      MKRKMMSLLLALAVISGSSVYAKVIDVMSPNGAIKVSVDIKDRIYYSVSY 50
gi|29339180|gb|AAO76978.1|       MKKLTFLLLCVLCTLS--LQAQKQFTLASPDGNLKTTITIGDRLTYDITC 48
                                 **:  : ** .*..:*      * : : **:* :*.:: * **: *.::

gi|60494642|emb|CAH09443.1|      DNDQLLKDCYLNLQLQN-ETLGTNPHLRSTKRGTIDESVKREIPFKNAIV 99
gi|29339180|gb|AAO76978.1|       NGRQILTPSPISMTLDNGTVWGENAKLSGTSRKSVDEMIPSPF-YRASEL 97
                                 :. *:*. . :.: *:*  . * *.:* .*.* ::** :   : :: : :

gi|60494642|emb|CAH09443.1|      RNHCNTLRMNFSGNYAVEFRVFDNGIAYRFVTDKKGDNIVMGEDFAINFP 149
gi|29339180|gb|AAO76978.1|       RNHYNGLTLRFKKDWNVEFRAYNDGIAYRFVNQGKKPFRVVTEVSDYCFP 147
                                 *** * * :.*. :: ****.:::*******.: *    *: *     **

gi|60494642|emb|CAH09443.1|      TNYKAHL----SQPDG-----FKTSYECPYTHVDTEKYAATDRMSYLPVL 190
gi|29339180|gb|AAO76978.1|       SDMTASVPYVKSGKDGDYNSQFFNSFENTYTTDKLSKLNK-QRLMFLPLV 196
                                 :: .* :    *  **     * .*:* .**  . .*    :*: :**::

gi|60494642|emb|CAH09443.1|      IETDKAYKILISEADLSDYPCMFLKSTG----KNGMQSIFPKAPLAFGED 236
gi|29339180|gb|AAO76978.1|       VDAGDGVKVCITESDLENYPGLYLSASEGANRLSSMHAPYPKRTVQGGHN 246
                                 ::::... *: *:*:**.:** ::*.::      ..*:: :** .:  *.:

gi|60494642|emb|CAH09443.1|      GDRSLKITEEADYIAKTDGKRSFPWRMMVISKEDKELIENEMVYNLSAPC 286
gi|29339180|gb|AAO76978.1|       -QLQMLVKEHEDYIAKVDKPRNFPWRIAVVTTTDKDLAATNLSYLLGAPS 295
                                  : .: :.*. *****.*  *.****: *::. **:*  .:: * *.**.

gi|60494642|emb|CAH09443.1|      VLEDYSWIKPGQVSWEWWHDARLYGVDFRSGFNMDSYKYYIDFASKFGIP 336
gi|29339180|gb|AAO76978.1|       RMSDLSWIKPGKVAWDWWNDWNLDGVDFVTGVNNPTYKAYIDFASANGIE 345
                                  :.* ******:*:*:**:* .* **** :*.*  :** ******  **

gi|60494642|emb|CAH09443.1|      YIIMDEGWAKNTR-DPFTPNPTINLTELIKYGKDRNVKIVLWLPWLTVEN 385
gi|29339180|gb|AAO76978.1|       YVILDEGWAVNLQADLMQVVKEIDLKELVDYAASKNVGIILWAGYHAFER 395
                                 *:*:***** * : * :     *:*.**:.*. .:** *:**  : :.*.

gi|60494642|emb|CAH09443.1|      HFD-LFKTFADWGIAGVKIDFMDRSDQWMVNYYERVAKEAAKHKLFVDFH 434
gi|29339180|gb|AAO76978.1|       DMENVCRHYAEMGVKGFKVDFMDRDDQEMTAFNYRAAEMCAKYKLILDLH 445
                                 .:: : : :*: *: *.*:*****.** *. :  *.*: .**:**::*:*

gi|60494642|emb|CAH09443.1|      GAFKPAGLERKYPNVLSYEGVLGMEQGGNCKP-----ENSIYLPFMRNAV 479
gi|29339180|gb|AAO76978.1|       GTHKPAGLNRTYPNVLNFEGVNGLEQMKWSSPSVDQVKYDVMIPFIRQVS 495
                                 *:.*****:*.*****.:*** *:**   ..*     : .: :**:*:.

gi|60494642|emb|CAH09443.1|      GPMDFTPGSMISAQPEDNRSTRANAMGSGTRAFQMALFIIFESGLQMLAD 529
gi|29339180|gb|AAO76978.1|       GPMDYTQGAMRNASKGNYYPCYSEPMSQGTRCRQLALYVVFESPFNMLCD 545
                                 ****:* *:* .*.  :  .  ::.*..***. *:**:::*** ::**.*

gi|60494642|emb|CAH09443.1|      NPVYYYRELPCTEFITSVPVTWDETKVLYAKVGEAVVVAKRKGEQWFIGG 579
gi|29339180|gb|AAO76978.1|       TPSNYMREPESTAFIAEIPTVWDESIVLDGKMGEYIVTARRKGDVWYVGG 595
                                 .*  * **  .* **:.:*..***: ** .*:** :*.*:***: *::**

gi|60494642|emb|CAH09443.1|      ITGNQPQNIEIDLGFIPAGQSFTLTSFEDGINADRQAMDYKKKESTVNNQ 629
gi|29339180|gb|AAO76978.1|       ITDWSARDIEVDCSFLGD-KSYHATLFKDGVNAHRAGRDYKCESFPIKKD 644
                                 **. ..::**:* .*:   :*:  * *:**:**.* . *** :. .::::

gi|60494642|emb|CAH09443.1|      TRMTLKMVRNGGWAGTIKMK 649
gi|29339180|gb|AAO76978.1|       GKLKVHLAPGGGFALKIK-- 662
                                  ::.:::. .**:* .**
```

Figure 48: Sequence alignment (of the CAH09443 amino acid residues and the AA076978 residues of *Bt*GH97b showing the conserved aspartic acid residue (emboldened and underlined) positioned at 415 and 404 respectively using ClustalW). (http://www.ebi.ac.uk/Tools/clustalw2/index.html).

Bibliography

Abbott, S. L. & Janda, J. M. (1997) '*Enterobacter cancerogenus* ("*Enterobacter taylorae*") - Infections associated with severe trauma or crush injuries', American Journal of Clinical Pathology, 107 (3), pp. 359-361.

Abdelal, A. T., Bibb, W. F. & Nainan, O. (1982) 'Carbamate Kinase from *Pseudomonas-Aeruginosa* - Purification, Characterization, Physiological-Role, and Regulation', Journal of Bacteriology, 151 (3), pp. 1411-1419.

Abergel, C., Bouveret, E., Claverie, J. M., Brown, K., Rigal, A., Lazdunski, C. & Benedetti, H. (1999) 'Structure of the *Escherichia coli* TolB protein determined by MAD methods at 1.95 angstrom resolution', Structure, 7 (10), pp. 1291-1300.

Anon (1994) 'Collaborative Computational Project Number 4', Acta Crystallographica Section D-Biological Crystallography, 50 (Pt 5), pp. 760-763.

Anon (1996) *Enzymatic assay of alpha-glucosidase.* Missouri: Sigma [Online]. Available at: www.sigmaaldrich.com (Accessed: 21/10/08).

Anon (2002) *Eliminate TFA and improve sensitivity of peptide analyses by LC/MS.* Bellefonte: Sigma-Aldrich Co. [Online]. Available at: http://www.sigmaaldrich.com/Graphics/Supelco/objects/11600/11547.pdf (Accessed: 25/06/2009).

Anon (2010) *Peptide mass fingerprint.* [Online]. Available at: http://www.matrixscience.com/help/pmf_help.html.

Anon (2009) *Dynamics and Inhibition of Class II fructose 1,6-bisphosphate aldolase.* Electronic thesis. University of Waterloo.

Barasa, N. W. (2008) Proteomic characterization of selenite resistance in a strain of *Enterobacter cloacae*. Youngstown State University.

Barkan, A., Klipcan, L., Ostersetzer, O., Kawamura, T., Asakura, Y. & Watkins, K. P. (2007) 'The CRM domain: An RNA binding module derived from an ancient ribosome-associated protein', Rna-a Publication of the Rna Society, 13 (1), pp. 55-64.

Barrios, A. F. G., Zuo, R. J., Ren, D. C. & Wood, T. K. (2006) 'Hha, YbaJ, and OmpA regulate *Escherichia coli* K12 biofilm formation and conjugation plasmids abolish motility', *Biotechnology and Bioengineering*, 93 (1), pp. 188-200.

Bartlett, D. H. & Welch, T. J. (1995) 'Omph Gene-Expression Is Regulated by Multiple Environmental Cues in Addition to High-Pressure in the Deep-Sea Bacterium Photobacterium Species Strain Ss9', *Journal of Bacteriology,* 177 (4), pp. 1008-1016.

Becker, P., Hufnagle, W., Peters, G. & Herrmann, M. (2001) 'Detection of differential gene expression in biofilm-forming versus planktonic populations of *Staphylococcus aureus* using micro-representational-difference analysis', *Applied and Environmental Microbiology,* 67 (7), pp. 2958-2965.

Beher, M. G., Schnaitman, C. A. & Pugsley, A. P. (1980) 'Major Heat-Modifiable Outer-Membrane Protein in Gram-Negative Bacteria - Comparison with the Ompa Protein of *Escherichia-Coli*', *Journal of Bacteriology,* 143 (2), pp. 906-913.

Bergey, D. H. & Holt, J. H. (1994) 'Bergey's manual of Determinative Bacteriology', in Bergey, D. H., Holt, J. H., Krieg, N. R. & Sneath, P. H. (eds.) Group 4- Gram negative aerobic/microaerophilic rods and cocci Group 5- Facultative anaerobic Gram negative rods. Lippincott Williams & Wilkins, pp. 102, 178.

Bernhardt, J., Weibezahn, J., Scharf, C. & Hecker, M. (2003) '*Bacillus subtilis* during feast and famine: Visualization of the overall regulation of protein synthesis

during glucose starvation by proteome analysis', *Genome Research,* 13 (2), pp. 224-237.

Bertrand, T., Briozzo, P., Assairi, L., Ofiteru, A., Bucurenci, N., Munier-Lehmann, H., Golinelli-Pimpaneau, B., Barzu, O. & Gilles, A. M. (2002) 'Sugar specificity of bacterial CMP kinases as revealed by crystal structures and mutagenesis of *Escherichia coli* enzyme', *Journal of Molecular Biology,* 315 (5), pp. 1099-1110.

Blakeley, P., Siepen, J.A., Lawless, C. & Hubbard, S.J . (2010). 'Investigating protein isoforms via proteomics: A feasible study.' *Proteomics* 10(6). pp. 1127-1140.

Blikslager, A. T., Moeser, A. J., Gookin, J. L., Jones, S. L. & Odle, J. (2007) 'Restoration of barrier function in injured intestinal mucosa', *Physiological Reviews*, 87 (2), pp. 545-564.

Bosch, M., Tarrago, R., Garrido, M. E., Campoy, S., de Henestrosa, A. R. F., de Rozas, A. M. P., Badiola, I. & Barbe, J. (2001) 'Expression of the *Pasteurella multocida* ompH gene is negatively regulated by the Fur protein', *FEMS Microbiology Letters,* 203 (1), pp. 35-40.

Boyd, E. F., Nelson, K., Wang, F. S., Whittam, T. S. & Selander, R. K. (1994) 'Molecular-Genetic Basis of Allelic Polymorphism in Malate-Dehydrogenase (Mdh) in Natural-Populations of *Escherichia-Coli* and *Salmonella-Enterica*', *Proceedings of the National Academy of Sciences of the United States of America,* 91 (4), pp. 1280-1284.

Brook, I. (2006) Bacteroides infection. [Online]. Available at: http://www.emedicine.com/med/topic2945.htm (Accessed: 15/12/2006).

Brotz-Oesterhelt, H., Bandow, J. E. & Labischinski, H. (2005) 'Bacterial proteomics and its role in antibacterial drug discovery', *Mass Spectrometry Reviews*, 24 (4), pp. 549-565.

Brown, T. A. (ed.) (1953) *Chapter 2: Vectors for gene cloning: Plasmids and Bacteriophages.* 3rd edn. London: Chapman and Hall, 1995, Gene cloning and DNA analysis: an introduction

Carlson, D. M. (1968) 'Structures and Immunochemical properties of oligosaccharides isolated from pig submaxillary mucins', *The Journal of Biological Chemistry*, 243 (3), pp. 616-626.

Carlsson, J., Nyberg, G. & Werthen, J. (1978) 'Hydrogen-Peroxide and Superoxide Radical Formation in Anaerobic Broth Media Exposed to Atmospheric Oxygen', *Applied and Environmental Microbiology,* 36 (2), pp. 223-229.

Cassataro, J., Velikovsky, C. A., Bruno, L., Estein, S. M., de la Barrera, S., Bowden, R., Fossati, C. A. & Giambartolomei, G. H. (2007) 'Improved immunogenicity of a vaccination regimen combining a DNA vaccine encoding *Brucella melitensis* outer membrane protein 31 (Omp31) and recombinant Omp31 boosting', *Clinical and Vaccine Immunology,* 14 (7), pp. 869-874.Brown, T. A. (ed.) (1953) *Chapter 2: Vectors for gene cloning: Plasmids and Bacteriophages.* 3rd edn. London: Chapman and Hall, 1995, Gene cloning and DNA analysis: an introduction

Cerdeno-Tarraga, A. M., Patrick, S., Crossman, L. C., Blakely, G., Abratt, V., Lennard, N., Poxton, I., Duerden, B., Harris, B., Quail, M. A., Barron, A., Clark, L., Corton, C., Doggett, J., Holden, M. T. G., Larke, N., Line, A., Lord, A., Norbertczak, H., Ormond, D., Price, C., Rabbinowitsch, E., Woodward, J., Barrell, B. & Parkhill, J. (2005) 'Extensive DNA inversions in the *B. fragilis* genome control variable gene expression', *Science,* 307 (5714), pp. 1463-1465.

Chang, D. E., Smalley, D. J. & Conway, T. (2002) 'Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model', *Molecular Microbiology,* 45 (2), pp. 289-306.

Chao, Y. P. & Liao, J. C. (1993) 'Alteration of Growth-Yield by Overexpression of Phosphoenolpyruvate Carboxylase and Phosphoenolpyruvate Carboxykinase in *Escherichia-Coli*', *Applied and Environmental Microbiology,* 59 (12), pp. 4261-4265.

Choe, L. H. & Lee, K. H. (2003) 'Quantitative and qualitative measure of intralaboratory two-dimensional protein gel reproducibility and the effects of sample preparation, sample load, and image analysis', *Electrophoresis,* 24 (19-20), pp. 3500-3507.

Corfield, A. P., Wagner, S. A., Clamp, J. R., Kriaris, M. S. & Hoskins, L. C. (1992) 'Mucin Degradation in the Human Colon - Production of Sialidase, Sialate O-Acetylesterase, N-Acetylneuraminate Lyase, Arylesterase, and Glycosulfatase Activities by Strains of Fecal Bacteria', *Infection and Immunity,* 60 (10), pp. 3971-3978.

Costerto.J.W., Ingram, J. M. & Cheng, K. J. (1974) 'Structure and Function of Cell-Envelope of Gram-Negative Bacteria', *Bacteriological Reviews,* 38 (1), pp. 87-110.

Cox, M. E. & Mangels, J. I. (1976) 'Improved Chamber for Isolation of Anaerobic Microorganisms', *Journal of Clinical Microbiology,* 4 (1), pp. 40-45.

Cuervo, P., Cupolillo, E., Britto, C., Gonzalez, L. J., Silva-Filho, F. C. E., Lopes, L. C., Domont, G. B. & De Jesus, J. B. (2008) 'Differential soluble protein expression between *Trichomonas vaginalis* isolates exhibiting low and high virulence phenotypes', *Journal of Proteomics,* 71 (1), pp. 109-122.

Davies, G. & Henrissat, B. (1995) 'Structures and Mechanisms of Glycosyl Hydrolases', Structure, 3 (9), pp. 853-859.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007) 'MolProbity: all-atom contacts and structure validation for

proteins and nucleic acids', *Nucleic Acids Research*, 35 (Web server issue), pp. 375-383.

Dekker, J., Rossen, J. W. A., Buller, H. A. & Einerhand, A. W. C. (2002) 'The MUC family: an obituary', *Trends in Biochemical Sciences*, 27 (3), pp. 126-131.

Derzelle, S., Bolotin, A., Mistou, M. Y. & Rul, F. (2005) 'Proteome analysis of *Streptococcus thermophilus* grown in milk reveals pyruvate formate-lyase as the major upregulated protein', *Applied and Environmental Microbiology,* 71 (12), pp. 8597-8605.

Dickey, R. S. & Zumoff, C. H. (1988) 'Emended Description of *Enterobacter-cancerogenus* Comb Nov (Formerly *Erwinia-cancerogena*)', International *Journal of Systematic Bacteriology*, 38 (4), pp. 371-374.

Diniz, C. G., Farias, L. M., Carvalho, M. A. R., Rocha, E. R. & Smith, C. J. (2004) 'Differential gene expression in a *Bacteroides fragilis* metronidazole-resistant mutant', *Journal of Antimicrobial Chemotherapy,* 54 (1), pp. 100-108.

Donnenberg, M. S., Zhang, H. Z. & Stone, K. D. (1997) 'Biogenesis of the bundle-forming pilus of enteropathogenic *Escherichia coli*: reconstitution of fimbriae in recombinant *E-coli* and role of DsbA in pilin stability - a review', *Gene,* 192 (1), pp. 33-38.

Donovan, R. S., Robinson, C. W. & Glick, B. R. (1996) 'Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter', *Journal of Industrial Microbiology,* 16 (3), pp. 145-154.

Edwards, R. & Read, P. N. (2000) 'Expression of the carbapenemase gene (cfiA) in *Bacteroides fragilis*', *Journal of Antimicrobial Chemotherapy*, 46 (6), pp. 1009-1012.

Eggers, C. T., Murray, L. A., Delmar, V. A., Day, A. G. & Craik, C. S. (2004) 'The periplasmic serine protease inhibitor ecotin protects bacteria against neutrophil elastase', *Biochemical Journal,* 379, pp. 107-118.

Emsley, P. & Cowtan, K. (2004) 'Coot: model-building tools for molecular graphics', Acta Crystallographica Section D-Biological Crystallography, 60, pp. 2126-2132.

Ernst, R. K., Dombroski, D. M. & Merrick, J. M. (1990) 'Anaerobiosis, Type-1 Fimbriae, and Growth-Phase Are Factors That Affect Invasion of Hep-2 Cells by *Salmonella-Typhimurium*', *Infection and Immunity,* 58 (6), pp. 2014-2016.

Eprintsev, A. T., Klimova, M. A., Shikhalieva, K. D. & Kompantseva, E. I. (2008) 'Isolation and purification of malate dehydrogenase isoforms from phototrophic purple bacteria *Rhodobacter sphaeroides* and *Rhodopseudomonas palustris*,' *Biology Bulletin,* 35(6), pp. 585-591.

Farmer, J. J., Fanning, G. R., Davis, B. R., Ohara, C. M., Riddle, C., Hickmanbrenner, F. W., Asbury, M. A., Lowery, V. A. & Brenner, D. J. (1985) '*Escherichia-Fergusonii* and *Enterobacter-Taylorae*, 2 New Species of Enterobacteriaceae Isolated from Clinical Specimens', J*ournal of Clinical Microbiology*, 21 (1), pp. 77-81.

Fasano, A. & Nataro, J. P. (2004) 'Intestinal epithelial tight junctions as targets for enteric bacteria-derived toxins', *Advanced Drug Delivery Reviews*, 56 (6), pp. 795-807.

Feniouk, B. A. & Junge, W. (2005) 'Regulation of the F0F1-ATP synthase: The conformation of subunit epsilon might be determined by directionality of subunit gamma rotation', *FEBS Letters,* 579 (23), pp. 5114-5118.

Ferguson, G. P., Totemeyer, S., MacLean, M. J. & Booth, I. R. (1998) 'Methylglyoxal production in bacteria: suicide or survival?', *Archives of Microbiology,* 170 (4), pp. 209-219.

Ferreiraa, E. O., Falcãoa, L. S., Vallima, D. C., Santosa, F. J., Andradeb, J. R. C., Andradeb, A. F. B., Vommaroc, R. C., Ferreiraa, M. C. S. & Domingues, R. M.

C. P. (2002) '*Bacteroides fragilis* adherence to Caco-2 cells', *Anaerobe*, 8 (6), pp. 307-314.

Fukushima, K., Ogawa, H., Takahashi, K., Naito, H., Funayama, Y., Kitayama, T., Yonezawa, H. & Sasaki, I. (2003) 'Non-pathogenic bacteria modulate colonic epithelial gene expression in germ-free mice', *Scandinavian Journal of Gastroenterology,* 38 (6), pp. 626-634.

Gietl, C. (1992) 'Malate-Dehydrogenase Isoenzymes - Cellular Locations and Role in the Flow of Metabolites between the Cytoplasm and Cell Organelles', *Biochimica Et Biophysica Acta,* 1100 (3), pp. 217-234.

Glaser, M., Nulty, W. & Vagelos, P. R. (1975) 'Role of Adenylate Kinase in Regulation of Macromolecular Biosynthesis in a Putative Mutant of *Escherichia-Coli* Defective in Membrane Phospholipid Biosynthesis', *Journal of Bacteriology,* 123 (1), pp. 128-136.

Gloster, T. M., Turkenburg, J. P., Potts, J. R., Henrissat, B. & Davies, G. J. (2008) 'Divergence of Catalytic Mechanism within a Glycosidase Family Provides Insight into Evolution of Carbohydrate Metabolism by Human Gut Flora', *Chemistry & Biology,* 59 (10), pp. 1058-1067.

Godoy, V. G., Dallas, M. M., Russo, T. A. & Malamy, M. H. (1993) 'A Role for *Bacteroides-Fragilis* Neuraminidase in Bacterial-Growth in 2 Model Systems', *Infection and Immunity*, 61 (10), pp. 4415-4426.

Gorg, A., Weiss, W. & Dunn, M. J. (2004) 'Current two-dimensional electrophoresis technology for proteomics', *Proteomics*, 4 (12), pp. 3665-3685.

Garazzino, S., Aprato, A., Maiello, A., Masse, A., Biasibetti, A., De Rosa, F. G. & Di Perri, G. (2005) 'Osteomyelitis caused by *Enterobacter cancerogenus* infection following a traumatic injury: Case report and review of the literature', *Journal of Clinical Microbiology*, 43 (3), pp. 1459-1461.

Grimont, P. A. D. & Ageron, E. (1989) '*Enterobacter-Cancerogenus* (Urosevic, 1966) Dickey and Zumoff 1988, a Senior Subjective Synonym of *Enterobacter-Taylorae* Farmer Et Al (1985)', *Research in Microbiology*, 140 (7), pp. 459-465.

Hegde, P. S., White, I. R. & Debouck, C. (2003) 'Interplay of transcriptomics and proteomics', *Current Opinion in Biotechnology*, 14 (6), pp. 647-651.

Heim, S., Lleo, M. D. M., Bonato, B., Guzman, C. A. & Canepari, P. (2002) 'The Viable but Nonculturable State and Starvation Are Different Stress Responses of *Enterococcus faecalis*, as Determined by Proteome Analysis ', *Journal of Bacteriology,* 184 (23), pp. 6739-6745.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P. & Davies, G. (1995) 'Conserved Catalytic Machinery and the Prediction of a Common Fold for Several Families of Glycosyl Hydrolases', *Proceedings of the National Academy of Sciences of the United States of America*, 92 (15), pp. 7090-7094.

Herrmann, K. M. (1995) 'The Shikimate Pathway as an Entry to Aromatic Secondary Metabolism', *Plant Physiology,* 107 (1), pp. 7-12.

Hogan, R. J., Mathews, S. A., Kutlin, A., Hammerschlag, M. R. & Timms, P. (2003) 'Differential expression of genes encoding membrane proteins between acute and continuous *Chlamydia pneumoniae* infections', *Microbial Pathogenesis,* 34 (1), pp. 11-16.

Holmgren, A., Johansson, C., Berndt, C., Lonn, M. E., Hudemann, C. & Lillig, C. H. (2005) 'Thiol redox control via thioredoxin and glutaredoxin systems', *Biochemical Society Transactions,* 33, pp. 1375-1377.

Hopley, L. & Schalkwyk, J. V. (2006) *Bacteroides fragilis*-beware of Bacteroides! [Online]. Available at:

http://www.anaesthetist.com/icu/infect/bacteria/anaerobe/Findex.htm#bfrag.htm (Accessed: 6-12-2006).

Hughes, C. V., Malki, G., Loo, C. Y., Tanner, A. C. R. & Ganeshkumar, N. (2003) 'Cloning and expression of alpha-D-glucosidase and N-acetyl-beta-glucosaminidase from the periodontal pathogen, Tannerella forsythensis (Bacteroides forsythus)', *Oral Microbiology and Immunology,* 18 (5), pp. 309-312.

Hwang, Y. W. & Miller, D. L. (1985) 'A Study of the Kinetic Mechanism of Elongation Factor-Ts', *Journal of Biological Chemistry,* 260 (21), pp. 1498-1502.

Hynes, M. J., Draht, O. W. & Davis, M. A. (2002) 'Regulation of the acuF gene, encoding phosphoenolpyruvate carboxykinase in the filamentous fungus *Aspergillus nidulans*', *Journal of Bacteriology,* 184 (1), pp. 183-190.

Ingrell, C. R. (2009) Challenges in biological interpretation of proteomics data.

Ito, K., Bassford, P. J. & Beckwith, J. (1981) 'Protein Localization in *Escherichia-Coli* - Is There a Common Step in the Secretion of Periplasmic and Outer-Membrane Proteins', *Cell,* 24 (3), pp. 707-717.

Janda, J. M. & Abbott, S. L. (2006) The enterobacteria. American Society for Biology [Online]. Available at: http://books.google.co.uk/books?hl=en&lr=&id=tjZT9_wRsXUC&oi=fnd&pg=PA1&dq=The+enterobacteria+book&ots=nVBGdM9Tjb&sig=HlQqvgJ3M7msVMyBuDCBl6i5oic (Accessed 19-07-2009)

Janosi, L., Shimizu, I. & Kaji, A. (1994) 'Ribosome Recycling Factor (Ribosome Releasing-Factor) Is Essential for Bacterial-Growth', *Proceedings of the National Academy of Sciences of the United States of America,* 91 (10), pp. 4249-4253.

Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T. & Watson, J. T. (1987) 'Novel Fragmentation Process of Peptides by Collision-Induced Decomposition in a Tandem Mass-Spectrometer - Differentiation of Leucine and Isoleucine', *Analytical Chemistry*, 59 (21), pp. 2621-2625.

Kabir, M. M. & Shimizu, K. (2003) 'Fermentation characteristics and protein expression patterns in a recombinant *Escherichia coli* mutant lacking phosphoglucose isomerase for poly(3-hydroxybutyrate) production', *Applied Microbiology and Biotechnology,* 62 (2-3), pp. 244-255.

Karlsson, N. G., Nordman, H., Karlsson, H., Carlstedt, I. & Hansson, G. C. (1997) 'Glycosylation differences between pig gastric mucin populations: A comparative study of the neutral oligosaccharides using mass spectrometry', *Biochemical Journal,* 326, pp. 911-917.

Katayama, T., Sakuma, A., Kimura, T., Makimura, Y., Hiratake, J., Sakata, K., Yamanoi, T., Kumagai, H. & Yamamoto, K. (2004) 'Molecular cloning and characterization of *Bifidobacterium bifidum* 1,2-alpha-L-fucosidase (AfcA), a novel inverting glycosidase (Glycoside hydrolase family 95)', *Journal of Bacteriology,* 186 (15), pp. 4885-4893.

Keller, R., Pedroso, M. Z., Ritchmann, R. & Silva, R. M. (1998) 'Occurence of virulence-associated properties in *E. cloacae*', *Infection and Immunity,* 66 (2), pp. 645-649.

Kim, K. J. (2001) 'Regulation of 3-deoxy-D-arabinoheptulosonate-7-phosphate (DAHP) synthase of *Bacillus sp* B-6 producing phenazine-1-carboxylic acid', *Journal of Biochemistry and Molecular Biology,* 34 (4), pp. 299-304.

Kim, S. J., Han, Y. H., Kim, I. H. & Kim, H. K. (1999) 'Involvement of ArcA and Fnr in expression of *Escherichia coli* thiol peroxidase gene', *Iubmb Life,* 48 (2), pp. 215-218.

Kitamura, M., Okuyama, M., Tanzawa, F., Mori, H., Kitago, Y., Watanabe, N., Kimura, A., Tanaka, I. & Yao, M. (2008) 'Structural and Functional Analysis of a Glycoside Hydrolase Family 97 Enzyme from *Bacteroides thetaiotaomicron*', *Journal of Biological Chemistry,* 283 (52), pp. 36328-36337.

Ko, K. S., Kuwahara, T., Lee, K. & Kook, Y. H. (2009) 'Population structure and distribution of virulence-related genes of *Bacteroides fragilis* isolates from Korea and Japan', *Diagnostic Microbiology and Infectious Disease,* 64 (3), pp. 340-343.

Komatsu, M., Carraway, C. A. C., Fregien, N. L. & Carraway, K. L. (1997) 'Reversible disruption of cell-matrix and cell-cell interactions by overexpression of sialomucin complex', *Journal of Biological Chemistry*, 272 (52), pp. 33245-33254.

Korithoski, B., Levesque, C. M. & Cvitkovitch, D. G. (2007) 'Involvement of the detoxifying enzyme lactoylglutathione lyase in *Streptococcus mutans* aciduricity', *Journal of Bacteriology,* 189 (21), pp. 7586-7592.

Kovacikova, G., Lin, W. & Skorupski, K. (2005) 'Dual regulation of genes involved in acetoin biosynthesis and motility/biofilm formation by the virulence activator AphA and the acetate-responsive LysR-type regulator AlsR in Vibrio cholerae', *Molecular Microbiology,* 57 (2), pp. 420-433.

Landry, S. J., Taher, A., Georgopoulos, C. & vanderVies, S. M. (1996) 'Interplay of structure and disorder in cochaperonin mobile loops', *Proceedings of the National Academy of Sciences of the United States of America,* 93 (21), pp. 11622-11627.

Laux, D. C., Cohen, P. S. & Conway, T. (eds.) (2005) Role of the mucus layer in bacterial colonization of the intestine- Chapter 15. Washington: ASM Press, Colonization of mucosal surfaces

Lazar, S. W. & Kolter, R. (1996) 'SurA assists the folding of *Escherichia coli* outer membrane proteins', *Journal of Bacteriology,* 178 (6), pp. 1770-1773.

Lazzaroni, J. C., Dubuisson, J. F. & Vianney, A. (2002) 'The Tol proteins of *Escherichia coli* and their involvement in the translocation of group A colicins', *Biochimie,* 84 (5-6), pp. 391-397.

Lazzaroni, J. C., Germon, P., Ray, M. C. & Vianney, A. (1999) 'The Tol proteins of *Escherichia coli* and their involvement in the uptake of biomolecules and outer membrane stability', *FEMS Microbiology Letters,* 177 (2), pp. 191-197.

Lemaire, M. & Wesolowski-Louvel, M. (2004) 'Enolase and glycolytic flux play a role in the regulation of the glucose pennease gene RAG1 of *Kluyveromyces lactis*', *Genetics,* 168 (2), pp. 723-731.

Len, A. C. L., Harty, D. W. S. & Jacques, N. A. (2004) 'Stress-responsive proteins are upregulated in *Streptococcus mutans* during acid tolerance', *Microbiology-Sgm,* 150, pp. 1339-1351.

Leppanen, V. M., Merckel, M. C., Ollis, D. L., Wong, K. K., Kozarich, J. W. & Goldman, A. (1999) 'Pyruvate formate lyase is structurally homologous to type I ribonucleotide reductase', *Structure,* 7 (7), pp. 733-744.

Leslie, A. G. W., *in Joint CCP4 and ESF-EACMB newsletter on protein crystallography* (Daresbury Laboratory, Warrington, UK., 1992), Vol. 26.

Liebler, D. (2002) Introduction to Proteomics- Tools for the New Biology. New Jersey: Humana Press, Protein digestion techniques, Protein expression profiling.

Lin, Y. F., Wu, M. S., Chang, C. C., Lin, S. W., Lin, J. T., Sun, Y. J., Chen, D. S. & Chow, L. P. (2006) 'Comparative immunoproteomics of identification and characterization of virulence factors from *Helicobacter pylori* related to gastric

cancer', *Molecular & Cellular Proteomics,* 5 (8), pp. 1484-1496Levett, P. N. (ed.) (1991) *Anaerobic Microbiology: A practical approach.* Practical approach series, Oxford: IRL at Oxford University Press.

Lipinska, B., King, J., Ang, D. & Georgopoulos, C. (1988) 'Sequence-Analysis and Transcriptional Regulation of the *Escherichia-Coli* Grpe Gene, Encoding a Heat-Shock Protein', *Nucleic Acids Research,* 16 (15), pp. 7545-7562.

Liptak, Z. (2005) Algorithmic and combinational questions in mass spectrometry and EST cludtering Bielfeld University.

Liu, K.-J. & Shih, N.-Y. (2007) 'The role of enolase in tissue invasion and metastasis of pathogens and tumour cells', *Journal of Cancer Molecules,* 3 (2), pp. 45-48.

Livingston, S. J., Kominos, S. D. & Yee, R. B. (1978) 'New Medium for Selection and Presumptive Identification of *Bacteroides-Fragilis* Group', *Journal of Clinical Microbiology*, 7 (5), pp. 448-453.

Lu, C. D., Kwon, D. H. & Abdelal, A. T. (1997) 'Identification of greA encoding a transcriptional elongation factor as a member of the carA-orf-carB-greA operon in *Pseudomonas aeruginosa* PAO1', *Journal of Bacteriology,* 179 (9), pp. 3043-3046.

Lugea, A., Salas, A., Casalot, J., Guarner, F. & Malagelada, J. R. (2000) 'Surface hydrophobicity of the rat colonic mucosa is a defensive barrier against macromolecules and toxins', *Gut*, 46 (4), pp. 515-521.

Luo, Y. G., Glisson, J. R., Jackwood, M. W., Hancock, R. E. W., Bains, M., Cheng, I. H. N. & Wang, C. L. (1997) 'Cloning and characterization of the major outer membrane protein gene (ompH) of *Pasteurella multocida* X-73', *Journal of Bacteriology,* 179 (24), pp. 7856-7864.

MacLean, M. J., Ness, L. S., Ferguson, G. P. & Booth, I. R. (1998) 'The role of glyoxalase I in the detoxification of methylglyoxal and in the activation of the KefB K+ efflux system in *Escherichia coli*', *Molecular Microbiology,* 27 (3), pp. 563-571.

Macy, J. M., Ljungdahl, L. G. & Gottschalk, G. (1978) 'Pathway of Succinate and Propionate Formation in *Bacteroides-Fragilis*', *Journal of Bacteriology,* 134 (1), pp. 84-91.

Malone, J. P., Radabaugh, M.R., Leimgruber, R.M. & Gerstenecker, G.G. (2001). 'Practical aspects of fluorescent staining for proteomic applications,' *Electrophoresis,* 22 (5), pp. 919-932.

Matsumura, Y., Takagi, M. & Imanaka, T. (1993) 'Regulation of *Escherichia-Coli* Superoxide-Dismutase Genes (sodA and sodB) by Oxygen', *Biotechnology Letters,* 15 (3), pp. 229-234.

Mcgrath, M. E., Gillmor, S. A. & Fletterick, R. J. (1995) 'Ecotin - Lessons on Survival in a Protease-Filled World', *Protein Science,* 4 (2), pp. 141-148.

Mcgrath, M. E., Hines, W. M., Sakanari, J. A., Fletterick, R. J. & Craik, C. S. (1991) 'The Sequence and Reactive Site of Ecotin - a General Inhibitor of Pancreatic Serine Proteases from *Escherichia-Coli*', *Journal of Biological Chemistry,* 266 (10), pp. 6620-6625.

Meslin, J. C., Fontaine, N. & Andrieux, C. (1999) 'Variation of mucin distribution in the rat intestine, caecum and colon: effect of the bacterial flora', Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology, 123 (3), pp. 235-239.

Messerschmidt, A., Huber, R., Poulos, T. & Wieghardt, K. (eds.) (2001) *Handbook of Metalloproteins.* Chichester: John Wiley and Sons Limited, iron superoxide dismutase.

Mikulskis, A., Aristarkhov, A. & Lin, E. C. C. (1997) 'Regulation of expression of the ethanol dehydrogenase gene (adhE) in *Escherichia coli* by catabolite repressor activator protein cra', *Journal of Bacteriology,* 179 (22), pp. 7129-7134.

Minarik, P., Tomaskova, N., Kollarova, M. & Antalik, M. (2002) 'Malate dehydrogenases - Structure and function', *General Physiology and Biophysics,* 21 (3), pp. 257-265.

Missall, T. A., Pusateri, M. E. & Lodge, J. K. (2004) 'Thiol peroxidase is critical for virulence and resistance to nitric oxide and peroxide in the fungal pathogen, *Cryptococcus neoformans*', *Molecular Microbiology,* 51 (5), pp. 1447-1458.

Mizrachi, I. (2002) *The NCBI Handbook.* NCBI [Online]. Available at: http://www.ncbi.nlm.nih.gov/books/bookres.fcgi/handbook/ch1.pdf (Accessed: 05-08-09).

Monahan, I. M., Betts, J., Banerjee, D. K. & Butcher, P. D. (2001) 'Differential expression of mycobacterial proteins following phagocytosis by macrophages', *Microbiology-Sgm,* 147, pp. 459-471.

Morona, R., Manning, P. A. & Reeves, P. (1983) 'Identification and Characterization of the Tolc Protein, an Outer-Membrane Protein from *Escherichia-Coli*', *Journal of Bacteriology,* 153 (2), pp. 693-699.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997) 'Refinement of macromolecular structures by the maximum-likelihood method', Acta Crystallographica Section D-Biological Crystallography, 53, pp. 240-255.

Nachin, L., Nannmark, U. & Nystrom, T. (2005) 'Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility', *Journal of Bacteriology,* 187 (18), pp. 6265-6272.

Nair, M. K. M. & Venkitanarayanan, K. (2007) 'Role of bacterial OmpA and host cytoskeleton in the invasion of human intestinal epithelial cells by *Enterobacter sakazakii*', *Pediatric Research*, 62 (6), pp. 664-669.

Nakano, M. M., Zhu, Y., Haga, K., Yoshikawa, H., Sonenshein, A. L. & Zuber, P. (1999) 'A mutation in the 3-phosphoglycerate kinase gene allows anaerobic growth of *Bacillus subtilis* in the absence of ResE kinase', *Journal of Bacteriology,* 181 (22), pp. 7087-7097.

Nakano, V., Gomes, T. A. T., Vieira, M. A. M., Ferreira, R. D. & Avila-Campos, M. J. (2007) 'bft gene subtyping in enterotoxigenic *Bacteroides fragilis* isolated from children with acute diarrhea', *Anaerobe*, 13 (1), pp. 1-5.

Namavar, F., Verweijvanvught, M. A. J. J. & Maclaren, D. M. (1991) 'A Study of the Candidate Virulence Factors of *Bacteroides-Fragilis*', *Journal of General Microbiology*, 137, pp. 1431-1435.

Naumoff, D. G. (2005) 'GH97 is a new family of glycoside hydrolases, which is related to the alpha-galactosidase superfamily', *BMC Genomics,* 6, pp. -.

Nicholson, W. L. (2008) 'The *Bacillus subtilis* ydjL (bdhA) Gene Encodes Acetoin Reductase/2,3-Butanediol Dehydrogenase', *Applied and Environmental Microbiology,* 74 (22), pp. 6832-6838.

Nikaido, H. (2003) 'Molecular basis of bacterial outer membrane permeability revisited', *Microbiology and Molecular Biology Reviews,* 67 (4), pp. 593-+.

Nystrom, T. & Neidhardt, F. C. (1994) 'Expression and Role of the Universal Stress Protein, Uspa, of *Escherichia-Coli* during Growth Arrest', *Molecular Microbiology,* 11 (3), pp. 537-544.

Okuyama, M., Kitamura, M., Hondoh, H., Kang, M. S., Mori, H., Kimura, A., Tanaka, I. & Yao, M. (2009) 'Catalytic Mechanism of Retaining alpha-Galactosidase

Belonging to Glycoside Hydrolase Family 97', *Journal of Molecular Biology,* 392 (5), pp. 1232-1241.

Overweg, K., Pericone, C. D., Verhoef, G. G. C., Weiser, J. N., Meiring, H. D., De Jong, A. P. J. M., De Groot, R. & Hermans, P. W. M. (2000) 'Differential protein expression in phenotypic variants of *Streptococcus pneumoniae*', *Infection and Immunity,* 68 (8), pp. 4604-4610.

Packschies, L., Theyssen, H., Buchberger, A., Bukau, B., Goody, R. S. & Reinstein, J. (1997) 'GrpE accelerates nucleotide exchange of the molecular chaperone DnaK with an associative displacement mechanism', *Biochemistry,* 36 (12), pp. 3417-3422.

Panina, E. M., Vitreschak, A. G., Mironov, A. A. & Gelfand, M. S. (2001) 'Regulation of aromatic amino acid biosynthesis in gamma-proteobacteria', *Journal of Molecular Microbiology and Biotechnology,* 3 (4), pp. 529-543.

Pantosti, A., Tzianabos, A. O., Reinap, B. G., Onderdonk, A. B. & Kasper, D. L. (1993) '*Bacteroides-Fragilis* Strains Express Multiple Capsular Polysaccharides', *Journal of Clinical Microbiology*, 31 (7), pp. 1850-1855.

Pape, T., Wintermeyer, W. & Rodnina, M. V. (1998) 'Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E-coli ribosome', *EMBO Journal,* 17 (24), pp. 7490-7497.

Park, S. J., Cotter, P. A. & Gunsalus, R. P. (1995) 'Regulation of Malate-Dehydrogenase (Mdh) Gene-Expression in *Escherichia-Coli* in Response to Oxygen, Carbon, and Heme Availability', *Journal of Bacteriology,* 177 (22), pp. 6652-6656.

Peptide fragmentation. (2007) [Online]. Available at: http://www.matrixscience.com/help/fragmentation_help.html. (Accessed: 18/03/09).

Pennington, S. R. & Dunn, M. J. (eds.) (2001) Proteomics: from protein sequence to function. Oxford: BIOS.

PerezVilar, J., Eckhardt, A. E. & Hill, R. L. (1996) 'Porcine submaxillary mucin forms disulfide-bonded dimers between its carboxyl-terminal domains', *Journal of Biological Chemistry,* 271 (16), pp. 9845-9850.

Persson, O., Valadi, A., Nystrom, T. & Farewell, A. (2007) 'Metabolic control of the *Escherichia coli* universal stress protein response through fructose-6-phosphate', *Molecular Microbiology,* 65 (4), pp. 968-978.

Pettersson, I. & Kurland, C. G. (1980) 'Ribosomal-Protein L7-L12 Is Required for Optimal Translation', *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences,* 77 (7), pp. 4007-4010.

Podolsky, D. K. (1985) 'Oligosaccharide Structures of Human Colonic Mucin', *Journal of Biological Chemistry,* 260 (14), pp. 8262-8271.

Podolsky, D. K. (1999) 'Mucosal Immunity and Inflammation V. Innate mechanisms of mucosal defense and repair: the best offense is a good defense', *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 277 (3), pp. G495-G499.

Podolsky, D. K. & Isselbacher, K. J. (1983) 'Composition of Human Colonic Mucin - Selective Alteration in Inflammatory Bowel-Disease', *Journal of Clinical Investigation*, 72 (1), pp. 142-153.

Pruzzo, C., Guzman, C. A. & Dainelli, B. (1989) 'Incidence of Hemagglutination Activity among Pathogenic and Non-Pathogenic *Bacteroides-Fragilis* Strains and Role of Capsule and Pili in Ha and Adherence', *FEMS Microbiology Letters*, 59 (1-2), pp. 113-118.

Pumbwe, L., Skilbeck, C. A. & Wexler, H. M. (2006) 'The *Bacteroides fragilis* cell envelope: Quarterback, linebacker, coach-or all three?', *Anaerobe,* 12 (5-6), pp. 211-220.

Radosevich, T. J., Reinhardt, T. A., Lippolis, J. D., Bannantine, J. P. & Stabel, J. R. (2007) 'Proteome and differential expression analysis of membrane and cytosolic proteins from *Mycobacterium avium subsp. paratuberculosis* strains K-10 and 187', *Journal of Bacteriology,* 189 (3), pp. 1109-1117.

Ranford, J. C., Coates, A. R. M. & Henderson, B. (2004) 'Chaperonins are cell-signalling proteins: the unfolding biology of molecular chaperones', *Expert Reviews in Molecular Medicine*, pp. 1-17.

Resch, A., Rosenstein, R., Nerz, C. & Gotz, F. (2005) 'Differential gene expression profiling of *Staphylococcus aureus* cultivated under biofilm and planktonic conditions', *Applied and Environmental Microbiology,* 71 (5), pp. 2663-2676.

Riedel, K. & Lehner, A. (2007) 'Identification of proteins involved in osmotic stress response in *Enterobacter sakazakii* by proteomics', *Proteomics,* 7 (8), pp. 1217-1231.

Rigal, A., Bouveret, E., Lloubes, R., Lazdunski, C. & Benedetti, H. (1997) 'The TolB protein interacts with the porins of *Escherichia coli*', *Journal of Bacteriology,* 179 (23), pp. 7274-7279.

Roberton, A. M. & Stanley, R. A. (1982) 'Invitro Utilization of Mucin by *Bacteroides-Fragilis*', *Applied and Environmental Microbiology,* 43 (2), pp. 325-330.

Robertson, K. P., Smith, C. J., Gough, A. M. & Rocha, E. R. (2006) 'Characterization of *Bacteroides fragilis* hemolysins and regulation and synergistic interactions of HlyA and HlyB', *Infection and Immunity*, 74 (4), pp. 2304-2316.

Rocha, E. R., Herren, C. D., Smalley, D. J. & Smith, C. J. (2003) 'The complex oxidative stress response of *Bacteroides fragilis*: the role of OxyR in control of gene expression', *Anaerobe,* 9 (4), pp. 165-173.

Rose, M. C. & Voynow, J. A. (2006) 'Respiratory tract mucin genes and mucin glycoproteins in health and disease', *Physiological Reviews*, 86 (1), pp. 245-278.

Rosner, J. L. & Martin, R. G. (2009) 'An Excretory Function for the *Escherichia coli* Outer Membrane Pore TolC: Upregulation of marA and soxS Transcription and Rob Activity Due to Metabolites Accumulated in tolC Mutants', *Journal of Bacteriology,* 191 (16), pp. 5283-5292.

Roszak, D. B. & Colwell, R. R. (1987) 'Survival Strategies of Bacteria in the Natural-Environment', *Microbiological Reviews,* 51 (3), pp. 365-379.

Rubinstien, E. M., Klevjeranderson, P., Smith, C. A., Drouin, M. T. & Patterson, J. E. (1993) '*Enterobacter-Taylorae*, a New Opportunistic Pathogen - Report of 4 Cases', J*ournal of Clinical Microbiology,* 31 (2), pp. 249-254.

Sabet, M., Lee, S. W., Nauman, R. K., Sims, T. & Um, H. S. (2003) 'The surface (S-) layer is a virulence factor of *Bacteroides forsythus*', *Microbiology-Sgm*, 149, pp. 3617-3627.

Salyers, A. A., Vercellotti, J. R., West, S. E. H. & Wilkins, T. D. (1977) 'Fermentation of Mucin and Plant Polysaccharides by Strains of Bacteroides from Human Colon', *Applied and Environmental Microbiology*, 33 (2), pp. 319-322.

Sawers, G. & Bock, A. (1988) 'Anaerobic Regulation of Pyruvate Formate-Lyase from *Escherichia-Coli* K-12', *Journal of Bacteriology,* 170 (11), pp. 5330-5336.

Schierack, P., Walk, N., Reiter, K., Weyrauch, K. D. & Wieler, L. H. (2007) 'Composition of intestinal Enterobacteriaceae populations of healthy domestic pigs', *Microbiology-Sgm*, 153, pp. 3830-3837.

Schirmer, T. (1998) 'General and specific porins from bacterial outer membranes', *Journal of Structural Biology,* 121 (2), pp. 101-109.

Schocke, L. & Weimer, P. J. (1997) 'Purification and characterization of phosphoenolpyruvate carboxykinase from the anaerobic ruminal bacterium *Ruminococcus flavefaciens*', *Archives of Microbiology,* 167 (5), pp. 289-294.

Serino, M., Luche, E., Chabo, C., Amar, J. & Burcelin, R. (2009) 'Intestinal microflora and metabolic diseases', Diabetes and Metabolism.

Sheldon, W. L., MacAuley, M. S., Taylor, E. J., Robinson, C. E., Charnock, S. J., Davies, G. J., Vocadlo, D. J. & Black, G. W. (2006) 'Functional analysis of a group A streptococcal glycoside hydrolase Spy1600 from family 84 reveals it is a beta-N-acetylglucosaminidase and not a hyaluronidase', *Biochemical Journal,* 399, pp. 241-247.

Shirai, T., Hung, V. S., Morinaka, K., Kobayashi, T. & Ito, S. (2008) 'Crystal structure of GH13 alpha-glucosidase GSJ from one of the deepest sea bacteria', *Proteins-Structure Function and Bioinformatics,* 73 (1), pp. 126-133.

Singamsetty, V. K., Wang, Y., Shimada, H. & Prasadarao, N. V. (2008) 'Outer membrane protein A expression in *Enterobacter sakazakii* is required to induce microtubule condensation in human brain microvascular endothelial cells for invasion', *Microbial Pathogenesis,* 45 (3), pp. 181-191.

Sinnott, M. L. (1990) 'Catalytic Mechanisms of Enzymatic Glycosyl Transfer', *Chemical Reviews*, 90 (7), pp. 1171-1202.

Smith, C. J., Rocha, E. R. & Paster, B. J. (eds.) (2006) The Prokaryotes: a handbook on the biology of bacteria. The medically important *Bacteroides* spp. in health and disease.

Solem, C., Koebmann, B. & Jensen, P. R. (2008) 'Control analysis of the role of triosephosphate isomerase in glucose metabolism in *Lactococcus lactis*', *Iet Systems Biology,* 2 (2), pp. 64-72.

Sowell, S. M., Norbeck, A. D., Lipton, M. S., Nicora, C. D., Callister, S. J., Smith, R. D., Barofsky, D. F. & Giovannoni, S. J. (2008) 'Proteomic analysis of stationary phase in the marine bacterium "*Candidatus Pelagibacter ubique*"', *Applied and Environmental Microbiology,* 74 (13), pp. 4091-4100.

Stein, J. H. (1998) Internal Medicine. Mosby A Times Mirror Company [Online]. Available at: http://books.google.co.uk/books?isbn=0815186983 (Accessed: 16-04-09).

Stellwag, E. J. & Hylemon, P. B. (1976) 'Purification and Characterization of Bile-Salt Hydrolase from *Bacteroides-Fragilis* Subsp Fragilis', *Biochimica Et Biophysica Acta*, 452 (1), pp. 165-176.

Stepanova, E., Lee, J., Ozerova, M., Semenova, E., Datsenko, K., Wanner, B. L., Severinov, K. & Borukhov, S. (2007) 'Analysis of promoter targets for *Escherichia coli* transcription elongation factor GreA in vivo and in vitro', *Journal of Bacteriology,* 189 (24), pp. 8772-8785.

Straus, D., Walter, W. & Gross, C. A. (1990) 'Dnak, Dnaj, and Grpe Heat-Shock Proteins Negatively Regulate Heat-Shock Gene-Expression by Controlling the Synthesis and Stability of Sigma-32', *Genes & Development,* 4 (12A), pp. 2202-2209.

Su, Z. D. & Honek, J. F. (2007) 'Emerging bacterial enzyme targets', *Current Opinion in Investigational Drugs*, 8 (2), pp. 140-149.

Studier, F. W. (2005) 'Protein production by auto-induction in high-density shaking cultures', *Protein Expression and Purification,* 41 (1), pp. 207-234.

Stulke, J. (2002) 'Control of transcription termination in bacteria by RNA-binding proteins that modulate RNA structures', *Archives of Microbiology,* 177 (6), pp. 433-440.

Takatsume, Y., Izawa, S. & Inoue, Y. (2005) 'Unique regulation of glyoxalase I activity during osmotic stress response in the fission yeast *Schizosaccharomyces pombe*: neither the mRNA nor the protein level of glyoxalase I increase under conditions that enhance its activity', *Archives of Microbiology,* 183 (3), pp. 224-227.

Terhorst, C., Moller, W., Laursen, R. & Wittmann.B (1973) 'Primary Structure of an Acidic Protein from 50-S Ribosomes of *Escherichia-Coli* Which Is Involved in Gtp Hydrolysis Dependent on Elongation Factors G and T', *European Journal of Biochemistry,* 34 (1), pp. 138-152.

Thauer, R. K., Jungermann, K. & Decker, K. (1977) 'Energy-Conservation in Chemotropic Anaerobic Bacteria', *Bacteriological Reviews,* 41 (1), pp. 100-180.

Twyman, R. (2003) Transcriptomics: The global study of gene expression at the RNA level. [Online]. Available at: http://genome.wellcome.ac.uk/doc_wtd020758.html. (Accessed: 12-02-09)

Tyers, M. & Mann, M. (2003) 'From genomics to proteomics', *Nature*, 422 (6928), pp. 193-197.

Van de Laar, F. A., Lucassen, P. L., Akkermans, R. P., Van de Lisdonk, F. H., Rutten, G. E. & Van Weel, C. (2005) 'alpha-Glucosidase inhibitors for patients with type 2 diabetes - Results from a cochrane systematic review and meta-analysis', *Diabetes Care,* 28 (1), pp. 154-163.

Varel, V. H. & Bryant, M. P. (1974) 'Nutritional Features of *Bacteroides-Fragilis* Subsp Fragilis', *Applied Microbiology,* 28 (2), pp. 251-257.

Virlogeux-Payant, I., Baucheron, S., Pelet, J., Trotereau, J., Bottreau, E., Velge, P. & Cloeckaert, A. (2008) 'TolC, but not AcrB, is involved in the invasiveness of multidrug-resistant *Salmonella enterica serovar Typhimurium* by increasing type III secretion system-1 expression', *International Journal of Medical Microbiology,* 298 (7-8), pp. 561-569.

Wagner, A. F. V., Schultz, S., Bomke, J., Pils, T., Lehmann, W. D. & Knappe, J. (2001) 'YfiD of Escherichia coli and Y061 of bacteriophage T4 as autonomous glycyl radical cofactors reconstituting the catalytic center of oxygen-fragmented pyruvate formate-lyase', *Biochemical and Biophysical Research Communications,* 285 (2), pp. 456-462.

Walter, S. & Buchner, J. (2002) 'Molecular chaperones - Cellular machines for protein folding', *Angewandte Chemie-International Edition,* 41 (7), pp. 1098-1113.

Weber, A. & Jung, K. (2002) 'Profiling early osmostress-dependent gene expression in *Escherichia coli* using DNA microarrays', *Journal of Bacteriology,* 184 (19), pp. 5502-5507.

Westblom, T. U. & Coggins, M. E. (1987) 'Osteomyelitis Caused by *Enterobacter-Taylorae*, Formerly Enteric Group-19', *Journal of Clinical Microbiology*, 25 (12), pp. 2432-2433.

Wexler, H. M. (2002) 'Outer-membrane pore-forming proteins in gram-negative anaerobic bacteria', *Clinical Infectious Diseases*, 35, pp. S65-S71.

Wheeler, R. (2009) 'Outer membrane proteomics of *P. multocida* isolates to identify outative host- specificity determinants ', *Bioscience Horizons,* 2 (1), pp. 1-12.

Wiggins, R., Hicks, S. J., Soothill, P. W., Millar, M. R. & Corfield, A. P. (2001) 'Mucinases and sialidases: their role in the pathogenesis of sexually transmitted infections in the female genital tract', *Sexually Transmitted Infections*, 77 (6), pp. 402-408.

Wilkins, J. C., Homer, K. A. & Beighton, D. (2002) 'Analysis of *Streptococcus mutans* proteins modulated by culture under acidic conditions', *Applied and Environmental Microbiology,* 68 (5), pp. 2382-2390.

Wilkins, J. C., Beighton, D. & Homer, K. A. (2003) 'Effect of acidic pH on expression of surface-associated proteins of *Streptococcus oralis*', *Applied and Environmental Microbiology,* 69 (9), pp. 5290-5296.

Williams, T. L., Monday, S. R., Edelson-Mammel, S., Buchanan, R. & Musser, S. M. (2005) 'A top-down proteomics approach for differentiating thermal resistant strains of *Enterobacter sakazakii*', *Proteomics,* 5 (16), pp. 4161-4169.

Willis, C. L., Cummings, J. H., Neale, G. & Gibson, G. R. (1996) 'In vitro effects of mucin fermentation on the growth of human colonic sulphate-reducing bacteria', *Anaerobe*, 2 (2), pp. 117-122.

Wilson, D. N., Schluenzen, F., Harms, J. M., Yoshida, T., Ohkubo, T., Albrecht, R., Buerger, J., Kobayashi, Y. & Fucini, P. (2005) 'X-ray crystallography study on ribosome recycling: the mechanism of binding and action of RRF on the 50S ribosomal subunit', *EMBO Journal,* 24 (2), pp. 251-260.

Wilson, M. (ed.) (2002) Bacterial disease mechanisms- An introduction to cellular microbiology, Cambridge: Cambridge University Press.

Wyborn, N. R., Messenger, S. L., Henderson, R. A., Sawers, G., Roberts, R. E., Attwood, M. M. & Green, J. (2002) 'Expression of the *Escherichia coli* yfiD gene responds to intracellular pH and reduces the accumulation of acidic metabolic end products', *Microbiology-Sgm,* 148, pp. 1015-1026.

Xiao, Z. J. & Xu, P. (2007) 'Acetoin metabolism in bacteria', *Critical Reviews in Microbiology,* 33 (2), pp. 127-140.

Xu, G. Q., Shin, S.B.Y. & Jaffrey, S.R. (2009). 'Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini.' *Proceedings of the National Academy of Sciences of the United States of America,* 106(46), pp. 19310-19315.

Xu, J., Mahowald, M. A., Ley, R. E., Lozupone, C. A., Hamady, M., Martens, E. C., Henrissat, B., Coutinho, P. M., Minx, P., Latreille, P., Cordum, H., Van Brunt, A., Kim, K., Fulton, R. S., Fulton, L. A., Clifton, S. W., Wilson, R. K., Knight, R. D. & Gordon, J. I. (2007) 'Evolution of symbiotic bacteria in the distal human intestine', *Plos Biology,* 5 (7), pp. 1574-1586.

Yakhnin, H., Yakhnin, A. V. & Babitzke, P. (2006) 'The trp RNA-binding attenuation protein (TRAP) of *Bacillus subtilis* regulates translation initiation of ycbK, a gene encoding a putative efflux protein, by blocking ribosome binding', *Molecular Microbiology,* 61 (5), pp. 1252-1266.

Yamaji, R., Fujita, K., Nakanishi, I., Nagao, K., Naito, M., Tsuruo, T., Inui, H. & Nakano, Y. (2004) 'Hypoxic up-regulation of triosephosphate isomerase expression in mouse brain capillary endothelial cells', *Archives of Biochemistry and Biophysics,* 423 (2), pp. 332-342.

Yu, J. & Kroll, J. S. (1999) 'DsbA: a protein-folding catalyst contributing to bacterial virulence', *Microbes and Infection,* 1 (14), pp. 1221-1228.

Yurewicz, E. C. & Moghissi, K. S. (1981) 'Purification of Human Mid-Cycle Cervical Mucin and Characterization of Its Oligosaccharides with Respect to Size, Composition, and Microheterogeneity', *Journal of Biological Chemistry*, 256 (22), pp. 1895-1904.

Zhang, J., Yasin, M., Carraway, C. A. C. & Carraway, K. L. (2006) 'MUC4 expression and localization in gastrointestinal tract and skin of human embryos', *Tissue & Cell*, 38 (4), pp. 271-275.

Zhang, M. (2006) Proteomic analysis of Streptococci. Northumbria University.

Zhang, M., McDonald, F. M., Sturrock, S. S., Charnock, S. J., Humphery-Smith, I., Black, G. W. (2007). 'Group A streptococcus cell-associated pathogenic proteins as revealed by growth in hyaluronic acid-enriched media,' *Proteomics*, 7(9), pp. 1379-1390.

Zivy, M. & de Vienne, D. (2000) 'Proteomics: a link between genomics, genetics and physiology', *Plant Molecular Biology,* 44 (5), pp. 575-580.

Zysk, G., Bongaerts, R. J. M., Ten Thoren, E., Bethe, G., Hakenbeck, R. & Heinz, H. P. (2000) 'Detection of 23 immunogenic pneumococcal proteins using convalescent-phase serum', *Infection and Immunity,* 68 (6), pp. 3740-3743.

APPENDICES

Appendix A- Chemicals, media and enzymes

This section of the Appendix contains a list of the chemicals, enzymes and media used in the study.

A1 Chemicals used in the study

Acros

Lactose

p-aminobenzoic acid

Cysteine

Imidazole

Iodoacetamide

Monosodium phosphate

Duchefa

Ethylenediaminetetraacetic acid disodium salt

Fisher BioReagents

Acrylamide/Bisacrylamide 37.5:1 40% solution

Brilliant blue G-250

Coomassie Blue R-250

Methanol

Fisher chemicals

Acetic acid, glacial

Acetone

Acetonitrile

Ammonium bicarbonate

Ammonium sulphate

Ethanol

Glycerol

D-(+)-Glucose

Ammonium per sulphate

N,N,N’,N’-tetra methylethylene diamine (TEMED)

Phosphoric acid

Potassium dihydrogen phosphate

Propan-2-ol

Dipotassium hydrogen phosphate

Sodium carbonate

Sodium hydroxide

Sodium tetraborate

Melford

Agarose (High gel strength)

Dithiothreitol

Glycine

N-[2-hydroxyethylpiperazine-N’-[2-ethanesulphonic acid] (HEPES)

Isopropyl-β-D-thiogalactopyranoside

Kanamycin

Sodium chloride

Sodium dodecyl sulphate

Tris [Hydroxymethyl] aminomethane (Tris-HCl)

Urea

Promega

Trypsin Gold

Riedel-deHaen

Hydrochloric acid

Sulphuric acid

Sigma

L-Aspartic acid

Ammonium sulphate

Bovine serum albumin (BSA)

Bromophenol blue

CHAPS

Citric acid

Dipotassium phosphate

Ethidium bromide

Ferrous sulphate

Magnesium sulphate

Magnesium chloride

β-mercaptoethanol

Pantothenic acid

Potassium chloride

Xylene cyanol

Potassium monophosphate monobasic

Phenol

Resazurin

Riboflavin

Sodium acetate

Sodium chloride

Sodium hydroxide

Sodium sulphite

Calcium chloride

Cobalt chloride

Cupric chloride

Manganese chloride

Nickel (II) chloride

Barium chloride

Strontium chloride

Zinc chloride

Bradfords reagent

Porcine gastric Mucin Type II

Porcine gastric mucin Type III

Media:

Oxoid

Agar (Bacteriological agar N° 1)

NZ amine (casein hydrolysate)

Tryptone

Yeast extract

Columbia Agar base

Anaerobic CO_2 generation compact [CO_2 Gen™ (Atmosphere Generation System)]

Difco

Casitone (Bactopeptone)

BactAlert™ Anaerobic Basal media

Hampton Research

Aquasil

Amino acids and vitamins used in media:

Acros	Sigma	Fisher Scientific	BDH	Aldrich
L-alanine	L-arginine	L-lysine	4-aminobenzoic acid	Isoleucine
L-aspartic acid	Ascorbic acid		L(-)- Proline	L-tryptophan
L-glutamic acid	L-asparagine			
Pyridoxine	Folic acid			
Riboflavin	Glutamine			
L-serine	D(+) Biotin			
L-threonine	L-histidine			
L-valine	L-leucine			
	Niacinamide			
	Nicotinic acid			
	L-phenylalanine			
	L-tyrosine			

Crystallisation chemicals:

Ammonium acetate- Fisher chemicals

Ammonium chloride- Fisher

Tri ammonium citrate- Fisher

Ammonium citrate dibasic- Sigma

Ammonium formate- Acros

Ammonium fluoride- Sigma

Ammonium dihydrogen orthophosphate- Fisher

Ammonium nitrate- Fisher

Ammonium sulphate- Sigma

Ammonium tartrate dibasic- Sigma

Bicine- Acros

Bis-Tris- Melford

Bis-Tris propane- Sigma

1, 4- Butanediol- Acros

Caesium chloride- Fisher

Calcium acetate (dried) - Fisher

Calcium chloride hexahydrate- Fisher

CAPS free acid- Melford

CHES- Melford

Citric acid- Fisher

Cobalt (II) chloride- Fisher

1, 4- Dioxane- Fisher

Diammonium hydrogen citrate- Fisher

HEPES, free acid- Melford

Hexadecyl trimethyl ammonium bromide- Sigma

1, 6- Hexanediol- Acros, Fisher

Jeffamine- Fluka

Imidazole (GPG) – Fisher

Iron (II) chloride- Fisher

DL-Malic acid- Fisher

Tertiary butyl alcohol- Fisher

Barbitone sodium- BDH

Barium chloride- Sigma

Calcium sulphate- Fisher

Cadmium chloride- Acros

Ethylene glycol- Fisher

Formaldehyde (40%) – BDH

Formamide- BDH

Lithium acetate dehydrate- Acros

Lithium chloride anhydrous- Fisher

Lithium sulphate- Fisher

Magnesium acetate- Sigma

Magnesium chloride- Fisher

Magnesium formate dehydrate- Fluka

Magnesium nitrate- Fisher

Magnesium sulphate- Fisher

MES- Fisher

2-methyl-1, 2- pentanediol- Acros

Nickel (II) chloride- Fisone

Nickel chloride hexahydrate- sigma

Nickel sulphate- BDH

Nickel (II) sulphate hexahydrate- Sigma

Pentaerythritol ethoxylate- Aldrich

Phenylmethyl sulfonyl fluoride- Sigma

Poly (acrylic acid) – Acros

Polyethylene glycol- Fluka

Polyethylene glycol 4000 grade- Fisher

Polyethylene glycol 6000 grade- Fisher

Potassium dichromate-Melford

Potassium hydrogen tartrate- Fisher

Potassium iodide- Fisone

Sodium cacodylate trihydrate- Fisher

Sodium cyanoborohydride- Sigma

Thiomersal- BDH

Tri-lithium citrate tetrahydrate- Fluka

Sodium chloride- Fisher

Potassium thiocyanate- Fisher

Sodium acetate trihydrate- Fisher

Potassium-dihydrogen orthophosphate- Fisher

Potassium nitrate- Fisher

Sodium formate- Fisher

Potassium chloride- Fisher

Potassium iodide- Fisher

Sodium sulphate anhydrous- Fisher

Zinc acetate- Fisher

Sodium thiocyanate dehydrate- Fisher

Tri-sodium citrate- Fisher

Potassium sulphate- Fisher

Sodium selenite- Sigma

Polyethylenimine- Sigma

Zinc sulphate- Fisher

di- sodium tetraborate- Fisher

Succinic acid- Fisher

Potassium bromide- Fisher

Di-sodium hydrogen orthophosphate dodecahydrate- Fisher

Sodium nitrate- Fisher

Sodium bromide- Fisher

Sodium fluoride- Fisher

Zinc chloride- Fisher

Sodium succinate- Fisher

Tri-sodium citrate- Fisher

Potassium sodium tartrate- Fisher

Trimethylamine N-oxide dehydrate- Acros

Tri potassium citrate- Fisher

Tris Base Ultrapure- Melford

Sodium propionate- Fisher

Polyvinyl pyrrolidone K15- Fluka

Sodium dihydrogen orthophosphate dehydrate- Fisher

Sodium tartrate- Fisher

Potassium formate- Fluka

Potassium fluoride anhydrous- Fluka

Potassium acetate- Fisher

Sodium selenite- Fisher

Sodium acetate trihydrate- Fisher

Di-potassium hydrogen orthophosphate- Fisher

Polyethylene glycol 20000- Fluka

Polyethylene glycol 5000 monomethyl ether- Fluka

Propan-2-ol- BDH

Polypropylene glycol 2000 grade- Fisher

Porpane-1, 2- diol- Fisher

Propylene glycol 400- Fluka

Pentaerythritol propoxylate- Aldrich

Polyethylene glycol 8000- Fisher

Polyethylene glycol 400 grade- Fisher

Polyethylene glycol 600 grade- Fisher

Polyethylene glycol 1000 grade- Fisher

Polyethylene glycol- Sigma

Polyethylene glycol 3000- Fluka

2-Ethoxyethanol- Fisher

Polyethylene glycol 200 grade- Fisher

Polyethylene glycol 300- Sigma

Glycerol- Fisher

Polyethylene glycol 2000 monomethyl ether- Fluka.

A2 Recipe for buffers used in the cloning experiment

Taq reaction buffer (10 X)

Tris pH 8.4	200 mM
KCl	500 mM
$MgCl_2$	50 mM

T4 DNA polymerase reaction buffer (10 X)

Tris- HCl pH 8.8	67 mM
$MgCl_2$	6.7 mM
$(NH_4)_2SO_4$	16.6 mM
β mercaptoethanol	10 mM
EDTA	6.7 µM

BSA	170 µg/mL

KOD polymerase reaction buffer (10 X)

Tris HCl pH 7.5 at 25°C	20 mM
$MgCl_2$	8 mM
DTT	7.5 mM
BSA	50 µg/mL

NE Buffer 2 (1 X)

NaCl	50 mM
Tris-HCl pH 7.9 at 25°C	10 mM
$MgCl_2$	10 mM
DTT	1 mM

The buffer is supplied at a 10 X concentration.

Appendix B

Lab Equipment

Autoclaving

A benchtop Prestige® Medical 2100 Classic autoclave was used to sterilise solutions at 121°C, 32 lb/inch^2 pressure for 20 min.

pH meter

pH of solutions were adjusted to the required values using a Jenway Ion Meter 3340 calibrated with the standard buffer values of pH 4.0, 7.0 and 9.2.

Spectrophotometer

The Cecil 2000 Series Spectrophotometer was used to determine the cell density of cultures of bacteria at various stages of growth. The wavelength was set to 660 nm for bacterial cultures and 420 nm for determining the optical density of Bradford assay reaction mixtures.

The Hitach Fluorescence Spectrophotometer was used to determine the enzyme activity against methylumbelliferone substrates. The excitation wavelength was set at 360 nm and the emission wavelength was set at 460 nm.

Incubators

A Gallenkamp orbital shaker was used to grow liquid bacterial cultures and a Gallenkamp static incubator was used to grow bacteria on solid media. The..... anaerobic chamber was used to grow bacteria anaerobically at 37°C. The static incubator maintained a stable anaerobic environment by being flushed with a mixture of 10% carbon-di-oxide gas, 80% nitrogen gas and 10% hydrogen gas.

Freeze drier

The trypsinised protein samples containing digested peptides were lyophilised using the Christ® Alpha 1-2 freeze drier.

Centrifugation

The Sigma* 1-15 bench top microcentrifuge was used to centrifuge samples of upto 1.5 ml in volume. The Sigma 3K18C refrigerated bench top centrifuge was used to centrifuge larger volumes of cultures or solutions (>1.5 ml). The rotors used have been detailed as follows:

The rotor number 11133 was used to harvest cells at 4000 xg in clay universal holders (part No. 17049, 25 ml plastic universals were used), concentrate proteins at 4000 xg in 30 kDa cut off protein concentrators which were inserted into clay universal holders (part No. 17049). The rotor number 12158 was used to separate soluble cell extract from lysed cells at 14000 xg in 30 ml plastic centrifuge tubes (part No. 31190030). The rotor number 12131 was used to isolate DNA from the Qiagen kit and to extract proteins from lysed cells by centrifuging at 14000 xg in 1.5 ml microcentrifuge tubes.

Sonication

The MSE Soniprep 150-ultrasonication machine was used to lyse cells.

Protein purification

Large scale purification of proteins was carried out using an AKTA prime and the FPLC system which was controlled by Prime View Software.

PCR machine

Eppendorf MasterCycler™ PCR machine was used to perform all the PCR reactions.

Electroporator

A Bio-Rad Gene Pulser Xcell™ with an attached Bio-Rad Pc and CE module was used for the transformation of electrocompetent cells. The Bio-Rad ShockPod™ was used to place the cuvettes.

Agarose gel kits

Owl Scientific EASY-CAST electrophoresis system was used for the electrophoresis of DNA and the power supply was an E-C 570-90 E-C Apparatus Corporation power pack.

2D Clean up kit from Amersham Biosciences

The solutions used in the Clean up kit include the precipitant solution which was used to in solubilise proteins, the co-precipitant solution which worked with the precipitant solution in enhancing the precipitation of proteins, the wash buffer to get rid of non protein contaminants and the wash additive to facilitate rapid and complete resuspension of the proteins in the solution. Due to trade reasons, the company have not revealed the exact composition of these solutions.

SDS-PAGE gel kits for small gels

A Bio-Rad Mini-PROTEAN 3 Cell kit was used for the electrophoresis of proteins and the same E-C 570-90 E-C Apparatus corporation power pack was used.

SDS-PAGE kit for 2D analysis of large gels

Protean II XL 2-D cells and accessories were used to perform the electrophoresis. The kit included a tank and a lid, a central cooling core unit, casting stand, sandwich clamps, alignment card and combs. The accessories provided include a 4 mm xi clamp notch vs. 13 mm XL clamp notch, 19 mm xi spacer vs. 8 mm XL spacer, 181 mm xi core gasket vs. 198 mm XL core gasket and a 153 mm xi comb vs. 184 mm XL comb.

The molecular biology kits used in the study include the Qiagen Miniprep kit for the purification of recombinant plasmids, the Qiagen Maxiprep kit for the large scale preparation of cloning vectors, the Qiagen Qiaquick gel extraction kit for the purification of DNA from agarose gels and the Qiagen DNeasy kit for the purification of genomic DNA.

Immobiline DryStrip Kit

Immobiline^{TM} DryStrip gels of pH range 4-7 and 3-10 were ordered as a part of the electrophoresis kit. The other parts include the electrodes (anode and cathode), try and electrode holder, DryStrip aligners, IEF electrode strips and the Bio-Rad power pack 200 and Universal power supply pack for protein and nucleic acid electrophoresis.

Gel Documentation

The Quantity One^{TM} software in the Bio-Rad Gel Doc 2000 system was used to visualise the agarose gels. The Bio-Rad GS-800 densitometer was used to visualise SDS-PAGE gels. Hard copies of the gel were obtained using the Mitsubishi Video Copy Processor (Model P91) with Mitsubishi thermal paper (K65HM-$_{CE}$/High density type, 110 mm X 21 m).

Gel spot analysis

The spots on the gels were compared and analysed using the Advanced PDQuest™ 8.0 software in the Bio-Rad Gel Doc 2000 system.

LC-MS analysis

The Ultimate 3000 Dionex LC system was used to separate the peptide fragments before Mass spectrometric analysis. The LC system was equipped with a vaccum degasser, a UV cell compartment, flow manager, column compartment fitted with an oven maintained at 60°C and the autosampler system with a multiwell sample microtitre plate.

Mass spectrometric analysis

The peptides of interest were identified using the Bruker Daltonics ESI-ion trap Mass spectrometer equipped with the Esquire series, Hystar sample processing, Chromeleon LC system and the Biotools software for Mascot database searches.

Appendix C

Vectors used in the cloning experiment

The vector used in cloning was pET-YSBLIC

Construct = pET-YSBLIC

GAGGAG BseRI

GGCGCGCC AscI

CATATG NdeI

PreScission (3C) Protease
CTGGAAGTTCTGTTCCAGGGGCCC

Parent vector, Pet-28a containing 3C protease site

5′-AGATATACCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGAAGTTCTGTTCCAGGGGCCCCATATGGCTAGCATGACTGGTGGACAGCAAATG-3′
3′-TCTATATGGTACCCGTCGTCGGTAGTAGTAGTAGTAGTGTCGTCGCCGGACCTTCAAGACAAGGTCCCCGGGGTATACCGATCGTACTGACCACCTGTCGTTTAC-5′
ArgArgTyrMetGlySerSerHisHisHisHisHisHisSerSerGlyLeuGluValLeuPheGlnGlyProHisMetAlaSerMetThrGlyGlyGlnGlnMet

YSBL pET-28a LIC

5′-AGATATACCATGGGCAGCAGCCATCATCACCACCACCACAGGCGCGCCTTCTCCTCACTGTTCCAGGGGCCCCATATGGCTAGCATGACTGGTGGACAGCAAATG-3′
3′-TCTATATGGTACCCGTCGTCGGTAGTAGTAGTAGTAGTGTCCGCGCGGAAGAGGAGTGACAAGGTCCCCGGGGTATACCGATCGTACTGACCACCTGTCGTTTAC-5′
ArgArgTyrMetGlySerSerHisHisHisHisHisHis

Non Cleavable Tag

LIC site

9/10 additional N-terminal amino acids

BseRI digest

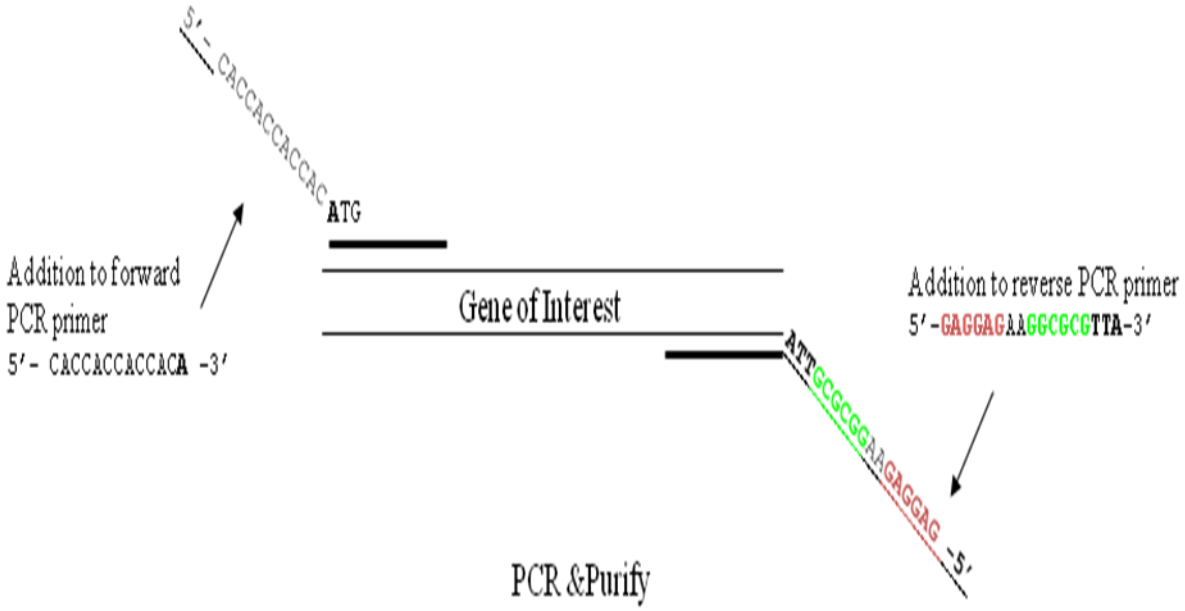
5′-GCAGCAGCCATCATCACCACCACCACAGG-3′
3′-CGTCGTCGGTAGTAGTGGTGGTGGTGT-5′

5′-CGCGCCTTCTCCTCACTGTTCCAGGGGCCCCATATGG-3′
3′-CCGCGCGGAAGAGGAGTGACAAGGTCCCCGGGGTATACC-5′

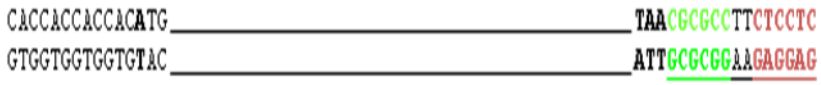
T4 pol/dTTP Treatment

GCAGCAGCAGCCATCAT-3′
CGTCGTCGTCGGTAGTAGTGGTGGTGGTGT -5′

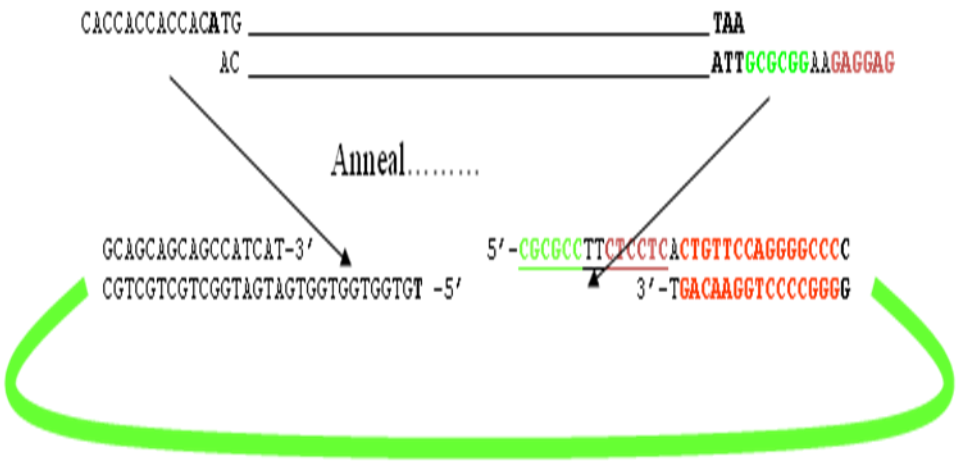
5′-CGCGCCTTCTCCTCACTGTTCCAGGGGCCCC
3′-TGACAAGGTCCCCGGGG



PCR &Purify



Treat with T4 pol and dATP



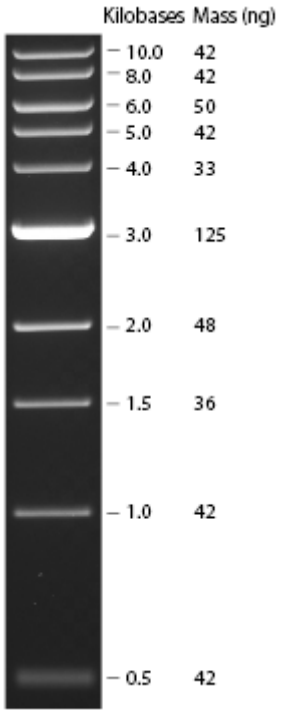
....then transform

Appendix D

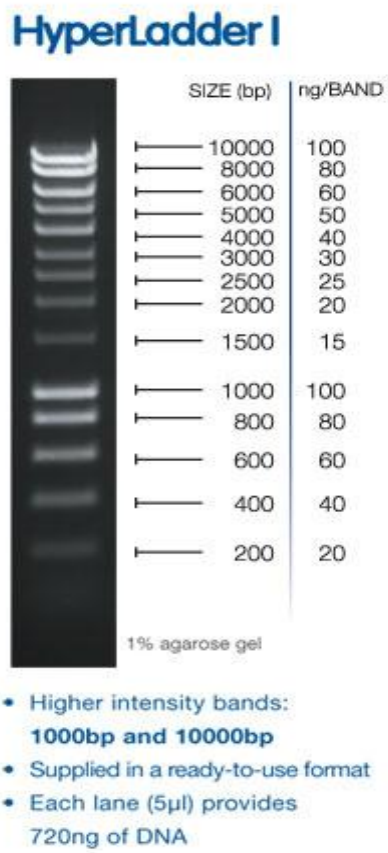
Protein and DNA size standards

Proteins	Molecular weight	High range (S8320)	Wide range (S8445)	Low range (M3913)
Myosin from rabbit muscle	200000	X	X	
B-galactosidase from *E. coli*	116000	X	X	
Phosphorylase b from rabbit muscle	97000	X	X	
Albumin, bovine serum	66000	X	X	X
Glutamic dehydrogenase from bovine liver	55000	X	X	
Ovalbumin from chicken egg	45000	X	X	X
Glyceraldehyde-3-phosphate dehydrogenase from rabbit muscle	36000	X	X	X
Carbonic anhydrase from bovine erythrocytes	29000		X	X
Trypsinogen from bovine pancreas	24000		X	X
Trypsin inhibitor from soyabean	20000		X	X
α-lactalbumin from bovine milk	14200		X	X
Aprotinin from bovine lung	6500		X	X

High and low molecular weight markers for SDS PAGE gels



NEB 1 kb Ladder used to determine the mass of the DNA fragments.



Bioline hyperladder I used as a size standard for agarose gel electrophoresis.

Appendix E

Growth curve data for *E. cancerogenus* and *B. fragilis*

Time in h	Absorbance at 600 nm- Experiment 1	Absorbance at 600 nm- Experiment 2	Absorbance at 600 nm- Experiment 3	Average absorbance at 600 nm
0	0	0	0	0
2	0.081	0.083	0.087	0.083667
3	0.416	0.58	0.55	0.515333
4	0.789	0.882	0.831	0.834
5	1.156	1.288	1.186	1.21
6	1.386	1.415	1.306	1.369
7	1.412	1.426	1.364	1.400667
8	1.434	1.445	1.474	1.451

Table 5: Growth of *E. cancerogenus* in control media without mucin

Time in h	Absorbance at 600 nm- Experiment 4	Absorbance at 600 nm- Experiment 5	Absorbance at 600 nm- Experiment 6	Average absorbance at 600 nm
0	0	0	0	0
2	0.053	0.076	0.108	0.079
3	0.21	0.256	0.257	0.241
4	0.931	0.794	0.786	0.837
5	1.264	1.132	1.188	1.194667
6	1.38	1.31	1.256	1.315333
7	1.392	1.370	1.376	1.379333
8	1.396	1.350	1.386	1.377333

Table 6: Growth of *E. cancerogenus* in media enriched with mucin Type II

Time in h	Absorbance at 600 nm- Experiment 7	Absorbance at 600 nm- Experiment 8	Absorbance at 600 nm- Experiment 9	Average absorbance at 600 nm
0	0	0	0	0
2	0.104	0.107	0.072	0.094333
3	0.31	0.292	0.280	0.294
4	0.901	0.897	0.807	0.868333
5	1.106	1.116	1.152	1.124667
6	1.186	1.242	1.264	1.230667
7	1.208	1.144	1.2	1.184
8	1.188	1.156	1.204	1.182667

Table 7: Growth of *E. cancerogenus* in media with mucin Type III

Time in h	Absorbance at 600 nm- Experiment 10	Absorbance at 600 nm- Experiment 11	Absorbance at 600 nm- Experiment 12	Average absorbance at 600 nm
0	0	0	0	0
4	0.067	0.087	0.059	0.071
6	0.112	0.1	0.158	0.123333
21	0.313	0.286	0.307	0.302
22	0.439	0.383	0.426	0.416
23	0.748	0.596	0.619	0.654333
24	1.004	0.909	0.93	0.947667
25	1.018	1.052	1.069	1.046333
26	1.394	1.298	1.243	1.311667
27	1.526	1.483	1.418	1.475667
28	1.556	1.552	1.512	1.54
29	1.488	1.418	1.385	1.430333
30	1.452	1.448	1.422	1.440667
31	1.475	1.466	1.486	1.475667
46	1.402	1.412	1.423	1.412333
47	1.328	1.396	1.404	1.376

Table 8: Growth of *B. fragilis* in control media without mucin

Time in h	Absorbance at 600 nm- Experiment 13	Absorbance at 600 nm- Experiment 14	Absorbance at 600 nm- Experiment 15	Average absorbance at 600 nm
0	0	0	0	0
4	0.012	0.018	0.015	0.015
6	0.05	0.041	0.045	0.045333
21	0.098	0.106	0.102	0.102
22	0.206	0.236	0.226	0.222667
23	0.34	0.358	0.36	0.352667
24	0.397	0.412	0.422	0.410333
25	0.532	0.622	0.603	0.585667
26	0.718	0.814	0.798	0.776667
27	0.922	1.039	0.956	0.972333
28	1.294	1.334	1.306	1.311333
29	1.433	1.461	1.457	1.450333
30	1.359	1.398	1.405	1.387333
31	1.382	1.368	1.393	1.381
46	1.401	1.387	1.384	1.390667
47	1.385	1.355	1.376	1.372

Table 9: Growth of *B. fragilis* in media with mucin Type II

Time in h	Absorbance at 600 nm- Experiment 16	Absorbance at 600 nm- Experiment 17	Absorbance at 600 nm- Experiment 18	Average absorbance at 600 nm
0	0	0	0	0
4	0.029	0.023	0.019	0.023667
6	0.063	0.052	0.058	0.057667
21	0.165	0.158	0.162	0.161667
22	0.384	0.4	0.391	0.391667
23	0.598	0.675	0.656	0.643
24	0.738	0.842	0.823	0.801
25	1.012	1.058	1.035	1.035
26	1.328	1.338	1.332	1.332667
27	1.438	1.463	1.457	1.452667
28	1.402	1.417	1.423	1.414
29	1.334	1.393	1.404	1.377
30	1.374	1.355	1.369	1.366
31	1.348	1.382	1.375	1.368333
46	1.368	1.376	1.36	1.368
47	1.284	1.318	1.307	1.303

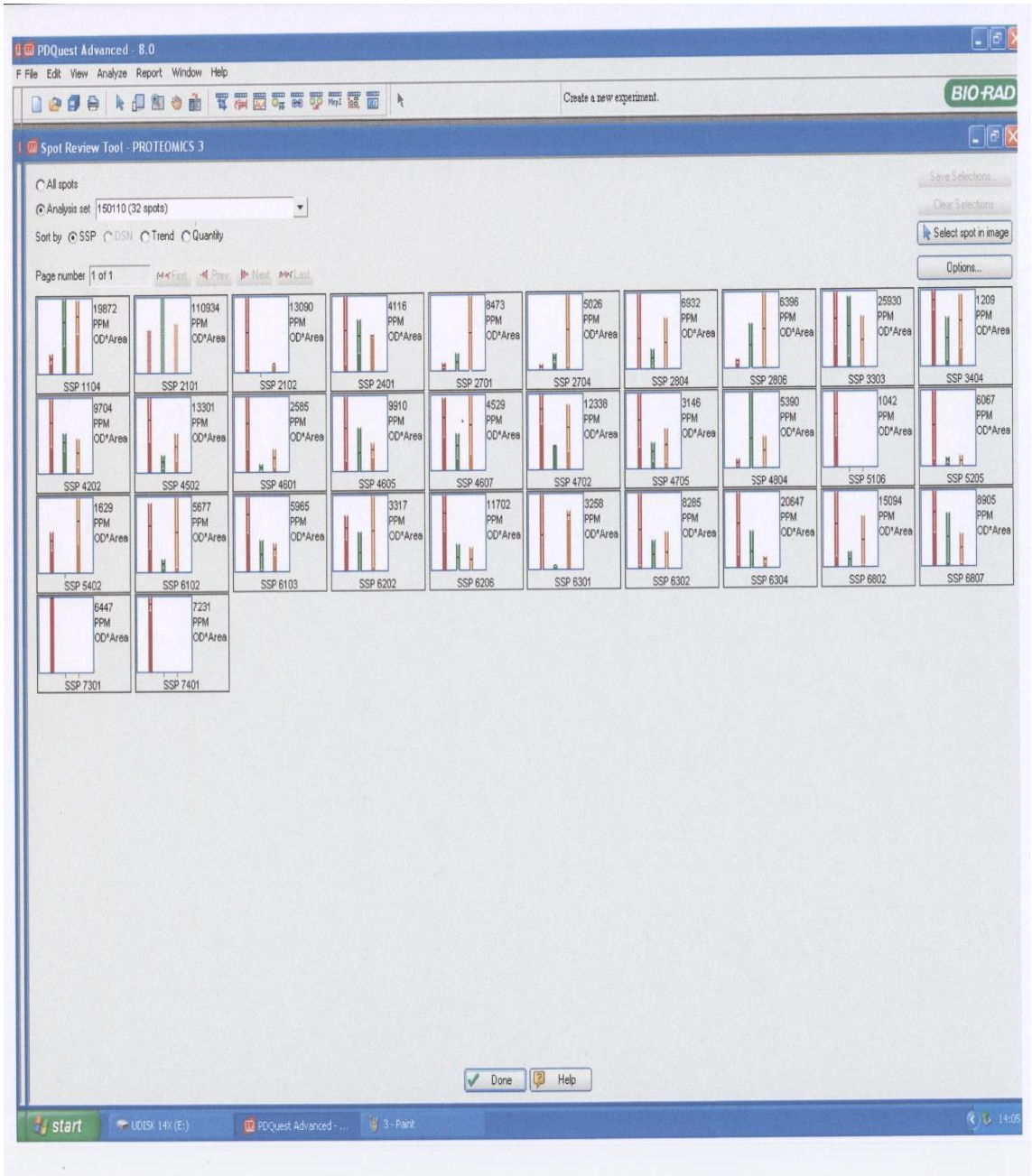
Table 10: Growth of *B. fragilis* in media with mucin Type III

E1 Resazurin test method

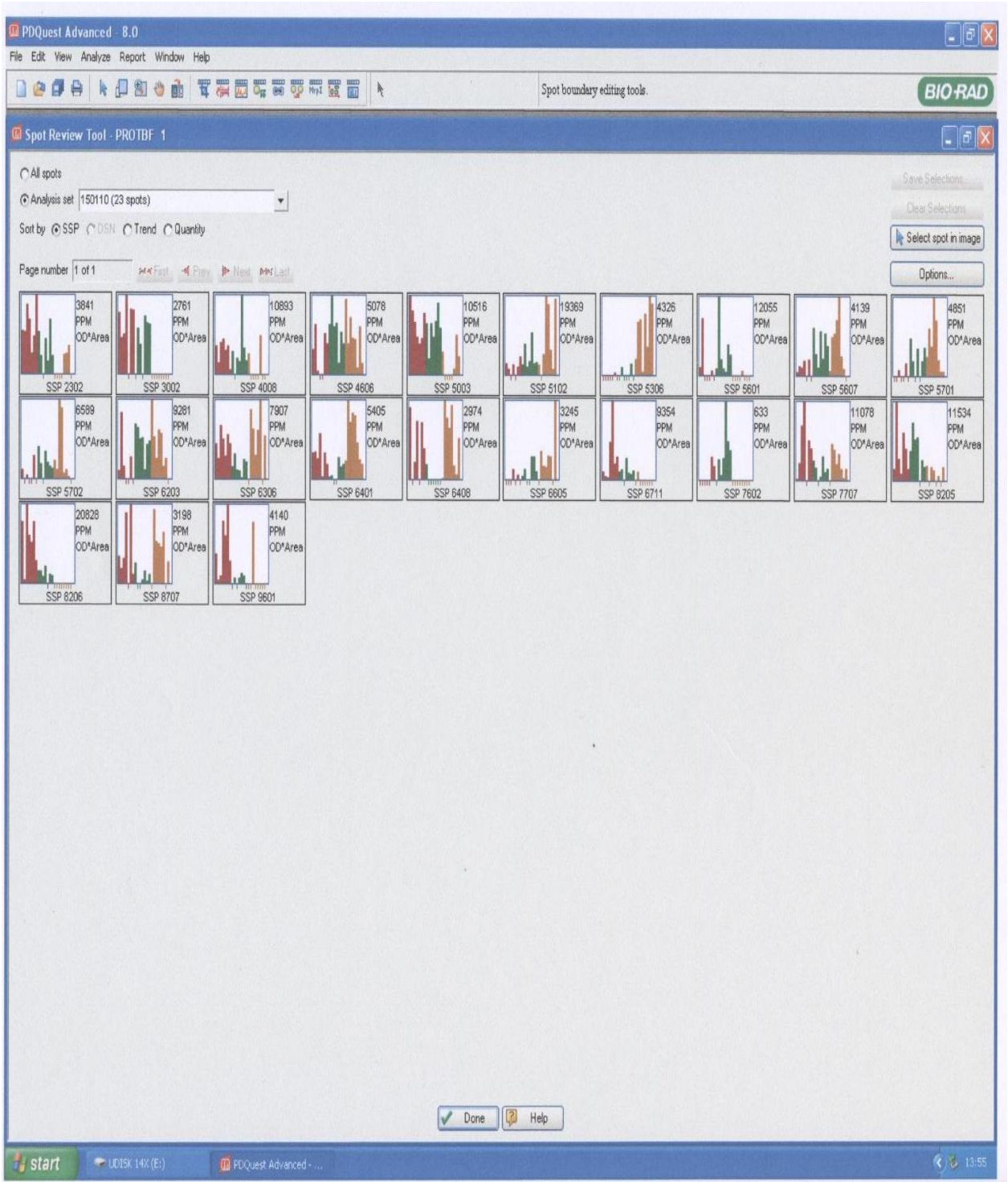
A 0.001% resazurin solution was prepared by dissolving 1 mg of resazurin and 2 g of cysteine-HCl in sterile 18.2 MΩ/cm making it up to 100 ml. The solution was aliquoted into 5 ml quantities and used as indicators. The resazurin strips or solutions were pink to purple in the presence of oxygen and gradually turn colourless when left in an environment lacking oxygen. This was normally used as a test to confirm the absence of oxygen in the bacterial growth environment.

Appendix F- Analysis data for 2DE gels

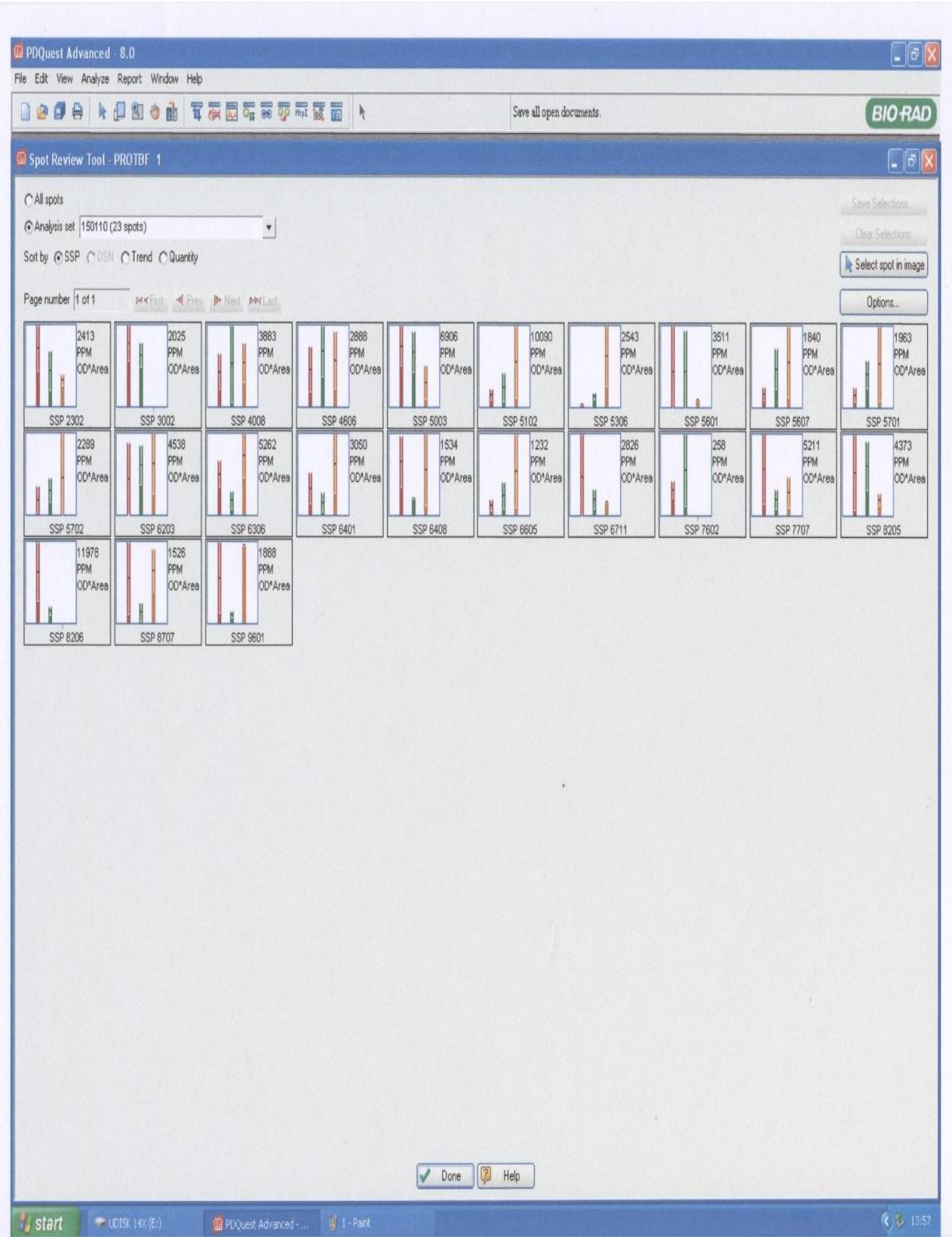
F1 Gel analysis data for *E. cancerogenus* and *B. fragilis* using Bio-Rad PDQuest software



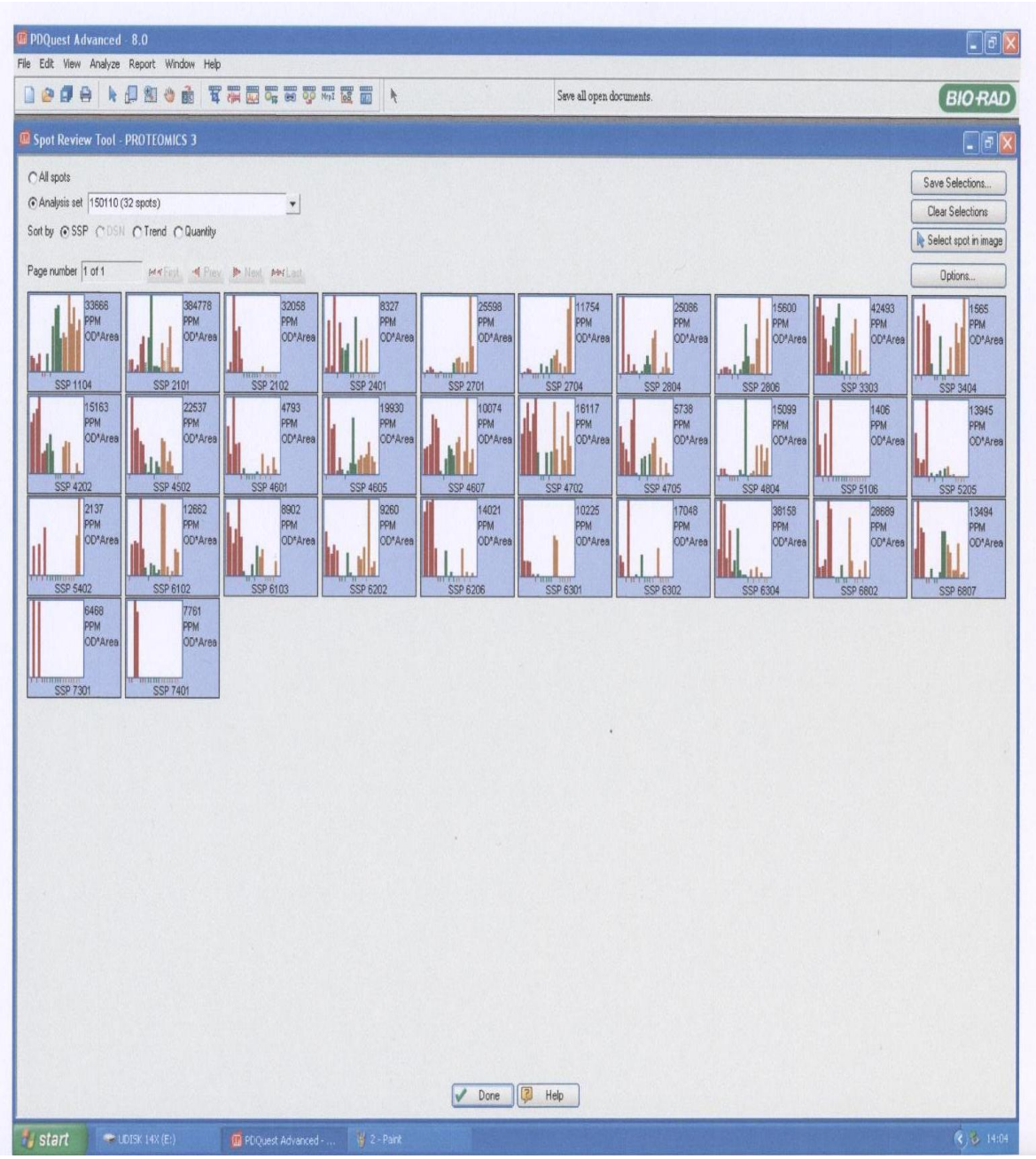
‘Protbf 1’ Average quantities of differentially expressed spots in three different growth conditions. Red bars indicate non-mucin media, Green bars- mucin Type II enriched media, Brown bars- mucin Type III enriched media.



‘Protbf 1’ Protein quantities in individual gels of differentially expressed spots in three different growth conditions. Red bars indicate non-mucin media, Green bars- mucin Type II enriched media, Brown bars- mucin Type III enriched media.



‘Proteomics 3’ Average quantities of differentially expressed spots in three different growth conditions. Red bars indicate non-mucin media, Green bars- mucin Type II enriched media, Brown bars- mucin Type III enriched media.



‘Proteomics 3’ Protein quantities in individual gels of differentially expressed spots in three different growth conditions. Red bars indicate non-mucin media, Green bars- mucin Type II enriched media, Brown bars- mucin Type III enriched media.

F2 Gel analysis data for *E. cancerogenus* and *B. fragilis* using Ludesi Redfin software

28/08/2009

Experiment overview

Experiment	PROTBF 1
Comparison	NON MUCIN MEDIA RD EC
Groups	MUCIN TYPE II MEDIA RD EC, NON MUCIN MEDIA RD EC, MUCIN TYPE III MEDIA RD EC

Experiment statistics

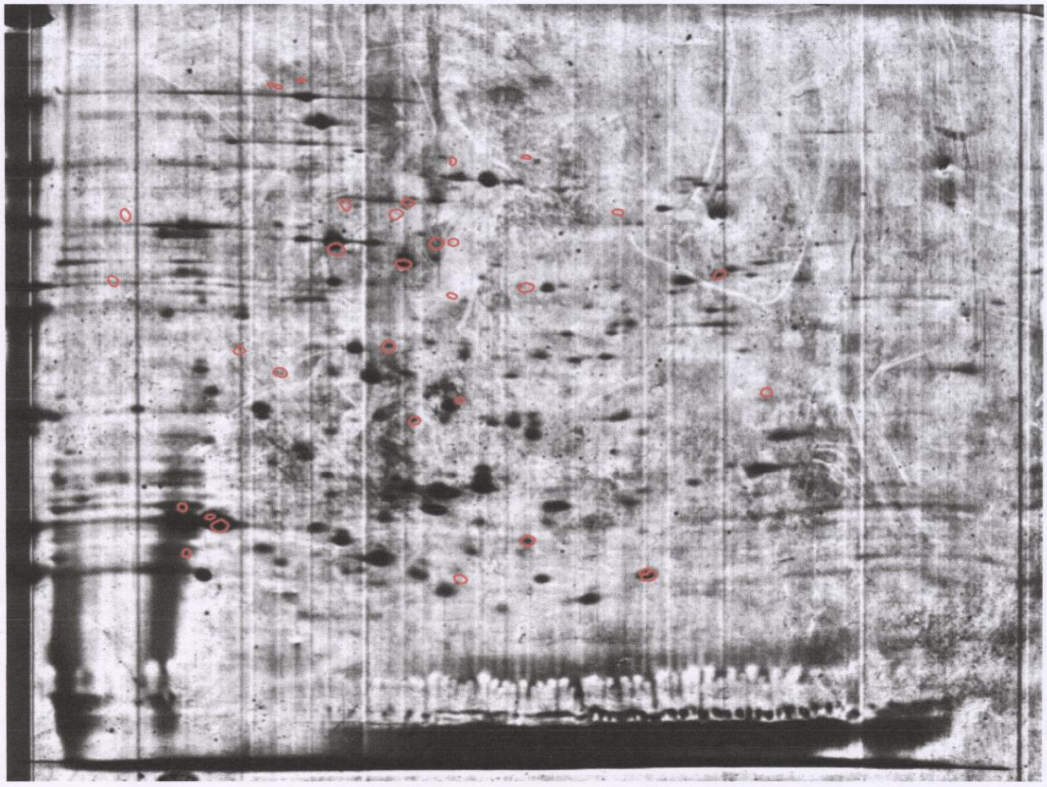
Statistical test	Anova
Applied filters	Anova < 0.05
	Presence >= 100
	Fold change > 1.5
	Location filter (top-left: 203, 238 to lower-right: 2323, 2853)
Number filtered spots	32

Group color legend

MUCIN TYPE II MEDIA RD EC NON MUCIN MEDIA RD EC MUCIN TYPE III MEDIA RD EC

Protein overview

Overview gel name: Scanned gel 10 (2) RD EC



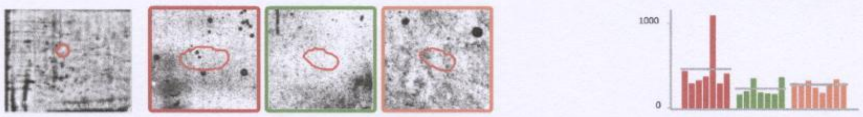
Overview

LUDESI

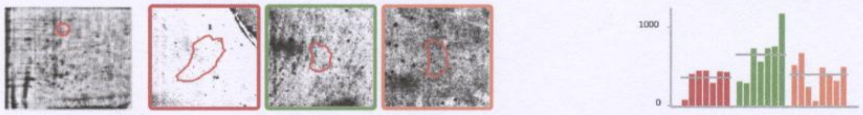
28/08/2009

Filtered spots

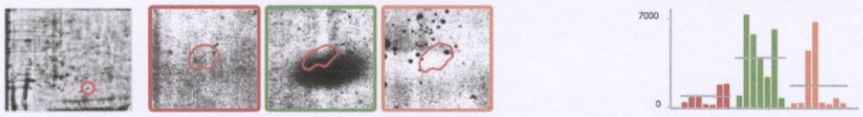
ID	619
Anova	0.0470
Presence	100 %
Fold	2.01
Volume	323



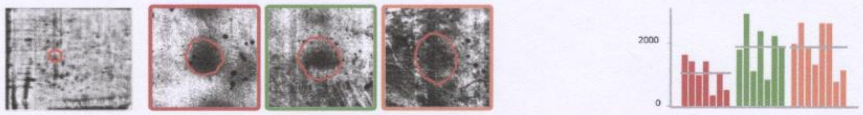
ID	733
Anova	0.0459
Presence	100 %
Fold	1.81
Volume	468



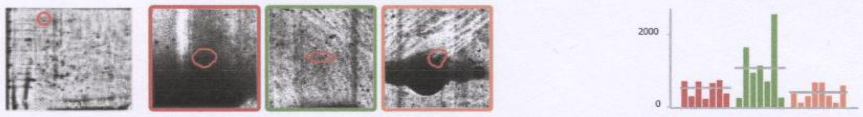
ID	127
Anova	0.0450
Presence	100 %
Fold	4.41
Volume	2140



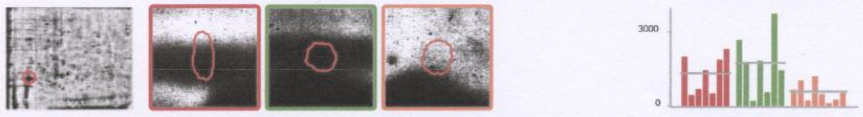
ID	320
Anova	0.0434
Presence	100 %
Fold	1.79
Volume	1608



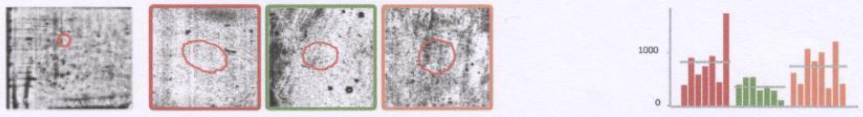
ID	681
Anova	0.0433
Presence	100 %
Fold	2.67
Volume	655



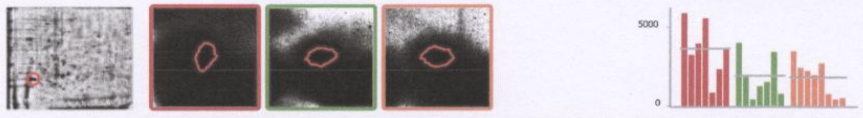
ID	583
Anova	0.0430
Presence	100 %
Fold	3.03
Volume	1182



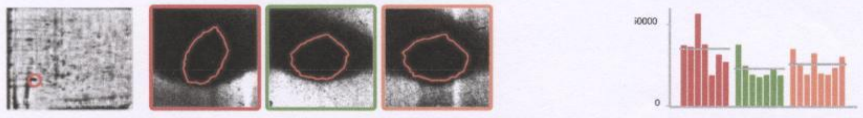
ID	695
Anova	0.0425
Presence	100 %
Fold	2.32
Volume	650



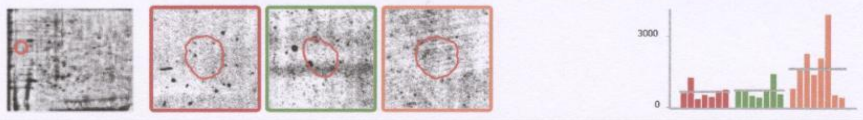
ID	636
Anova	0.0418
Presence	100 %
Fold	2.01
Volume	2436



ID	24
Anova	0.0354
Presence	100 %
Fold	1.54
Volume	27676



ID	445
Anova	0.0345
Presence	100 %
Fold	2.46
Volume	1025

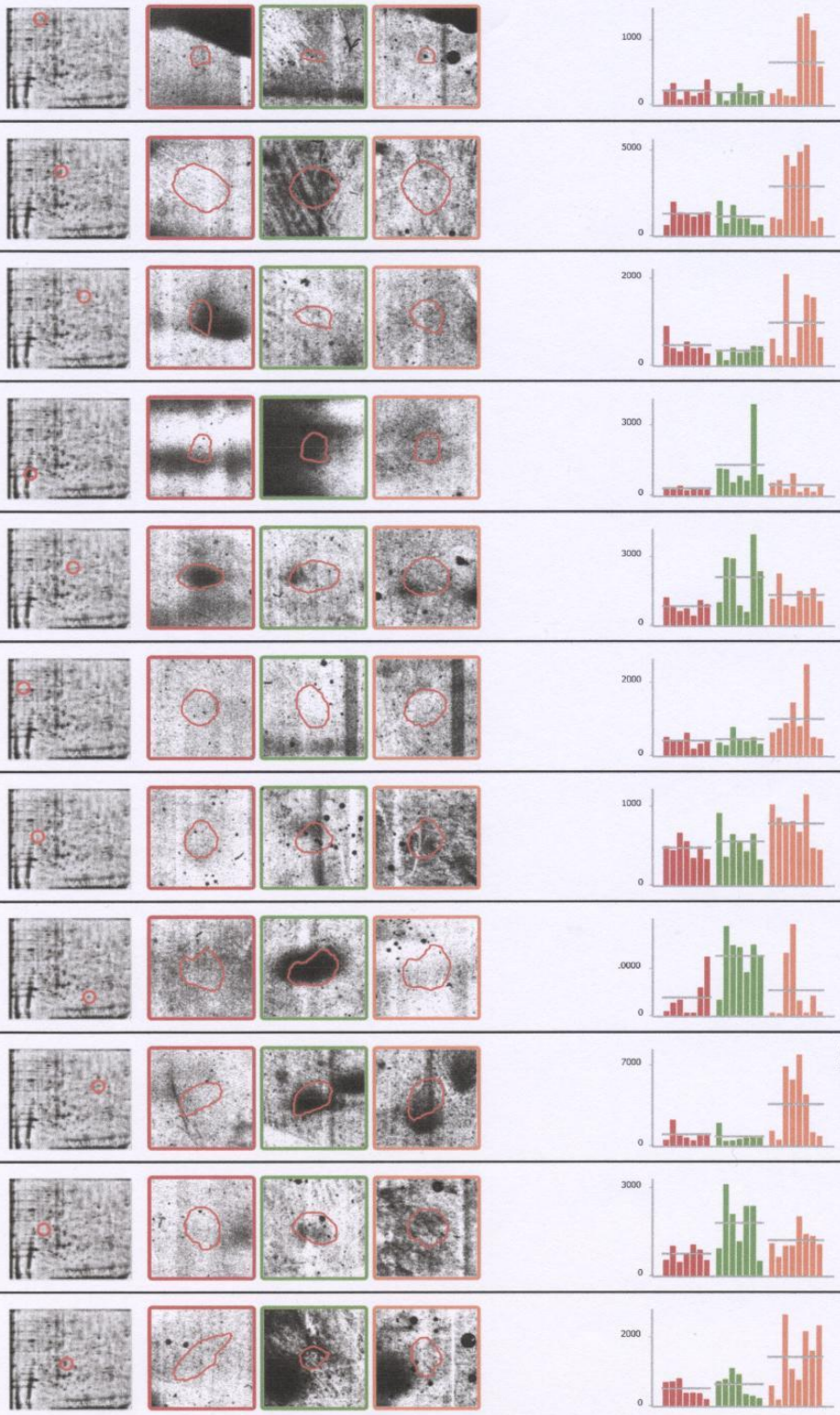


Item 1 to 10 of 32

LUDESI

28/08/2009

ID	672	
Anova	0.0324	
Presence	100 %	
Fold	3.44	
Volume	366	
ID	87	
Anova	0.0296	
Presence	100 %	
Fold	2.58	
Volume	1795	
ID	483	
Anova	0.0262	
Presence	100 %	
Fold	2.93	
Volume	609	
ID	551	
Anova	0.0250	
Presence	100 %	
Fold	4.36	
Volume	667	
ID	210	
Anova	0.0247	
Presence	100 %	
Fold	2.55	
Volume	1400	
ID	565	
Anova	0.0242	
Presence	100 %	
Fold	2.45	
Volume	635	
ID	684	
Anova	0.0204	
Presence	100 %	
Fold	1.64	
Volume	610	
ID	58	
Anova	0.0202	
Presence	100 %	
Fold	3.32	
Volume	7216	
ID	131	High fold increase Type III
Anova	0.0144	
Presence	100 %	
Fold	4.57	
Volume	1871	
ID	307	
Anova	0.0143	
Presence	100 %	
Fold	2.40	
Volume	1247	
ID	548	
Anova	0.0143	
Presence	100 %	
Fold	2.85	
Volume	871	

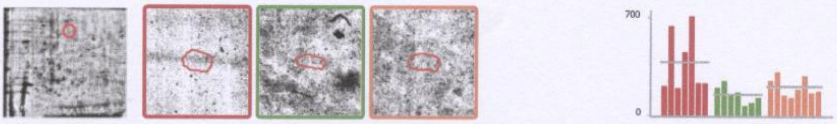


Item 11 to 21 of 32

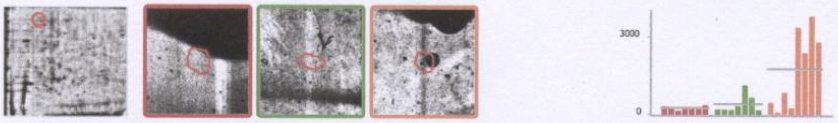
LUDESI

28/08/2009

ID	631
Anova	0.0116
Presence	100 %
Fold	2.55
Volume	245

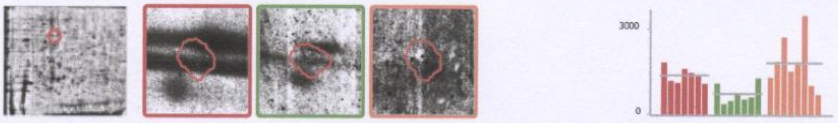


ID	130
Anova	0.0100
Presence	100 %
Fold	6.67
Volume	862

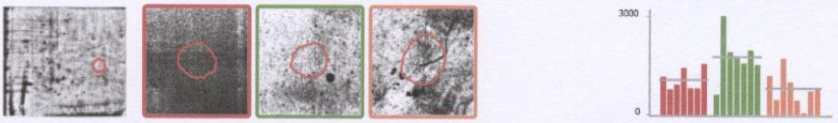


High fold increase in Mucin III

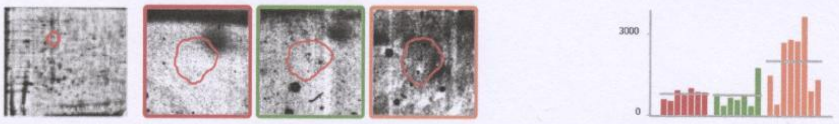
ID	297
Anova	0.0089
Presence	100 %
Fold	2.51
Volume	1325



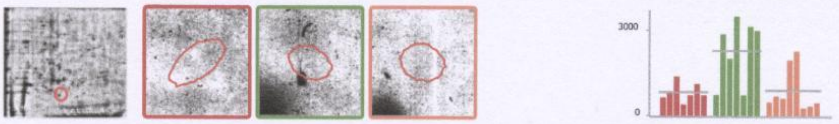
ID	384
Anova	0.0085
Presence	100 %
Fold	2.18
Volume	1205



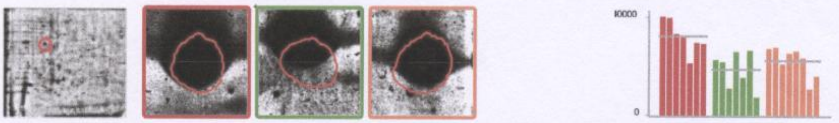
ID	137
Anova	0.0050
Presence	100 %
Fold	2.71
Volume	1208



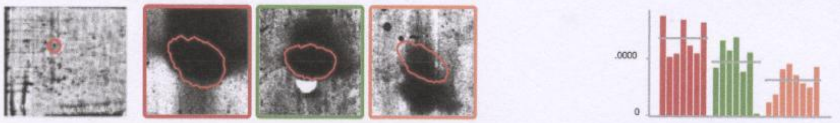
ID	351
Anova	0.0047
Presence	100 %
Fold	2.75
Volume	1312



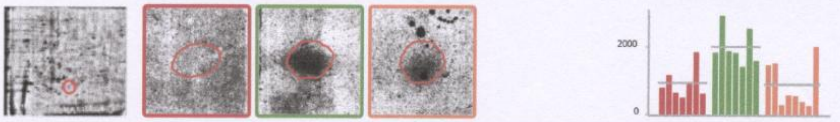
ID	1
Anova	0.0033
Presence	100 %
Fold	1.73
Volume	24393



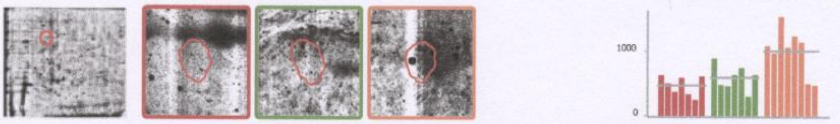
ID	48
Anova	0.0026
Presence	100 %
Fold	2.17
Volume	9679



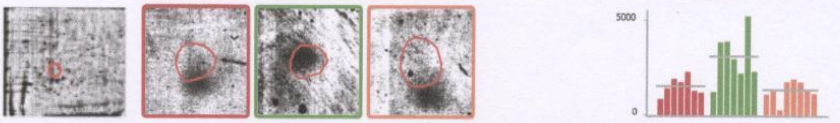
ID	437
Anova	0.0020
Presence	100 %
Fold	2.25
Volume	1262



ID	497
Anova	0.0020
Presence	100 %
Fold	2.13
Volume	705



ID	238
Anova	0.0016
Presence	100 %
Fold	2.33
Volume	1956



Item 22 to 32 of 32

LUDESI

04/01/2010

Experiment overview

Experiment	PROTEOMICS 3
Comparison	PROTEOMICS 3
Groups	NON MUCIN MEDIA RD, MUCIN TYPE II MEDIA RD, MUCIN TYPE III MEDIA RD

Experiment statistics

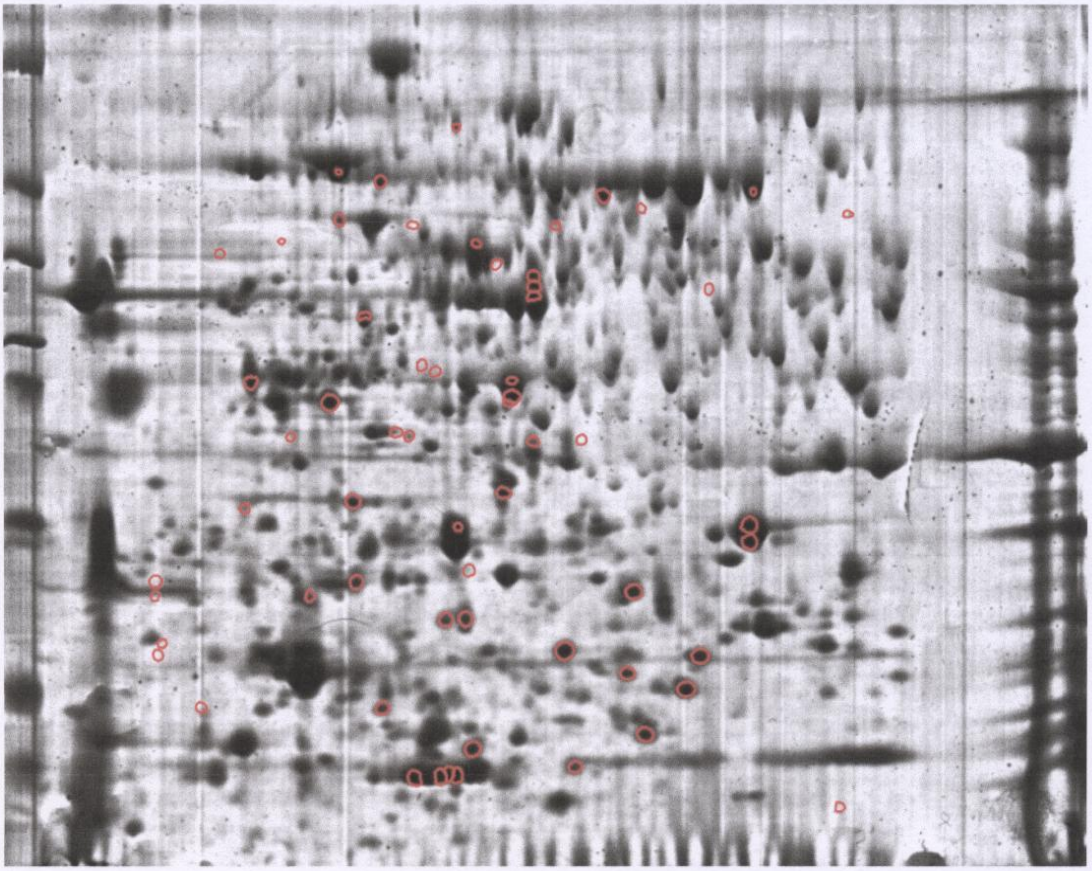
Statistical test	Anova
Applied filters	Anova < 0.05
	Presence >= 100
	Fold change > 1.5
	Location filter (top-left: 87, 195 to lower-right: 2272, 2755)
Number filtered spots	61

Group color legend

NON MUCIN MEDIA RD MUCIN TYPE II MEDIA RD MUCIN TYPE III MEDIA RD

Protein overview

Overview gel name: Staff 2008-10-24 14hr 21min (10) 4 RD



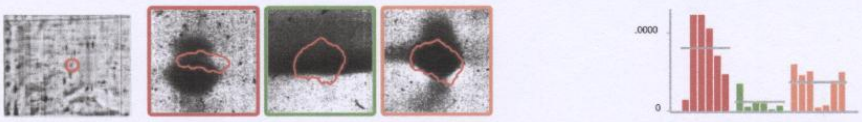
Overview

LUDESI

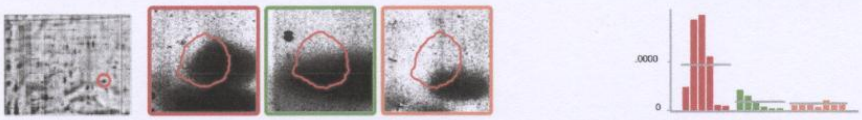
04/01/2010

Filtered spots

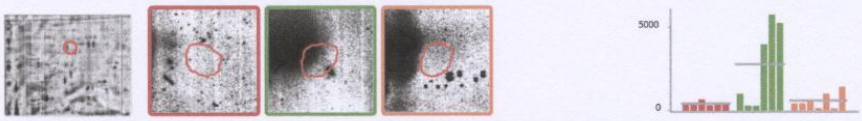
ID	10
Anova	0.0028
Presence	100 %
Fold	7.04
Volume	4270



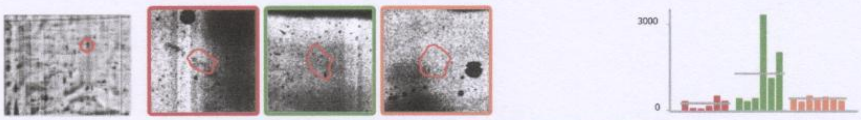
ID	17
Anova	0.0156
Presence	100 %
Fold	6.78
Volume	4047



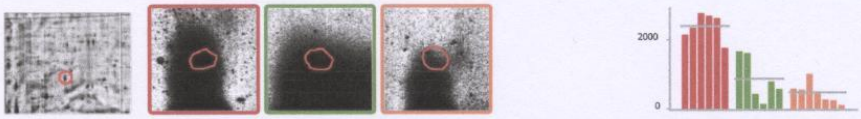
ID	138
Anova	0.0204
Presence	100 %
Fold	6.56
Volume	1230



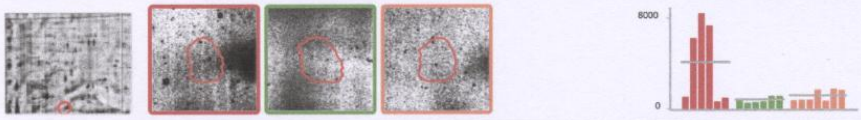
ID	8
Anova	0.0375
Presence	100 %
Fold	5.16
Volume	636



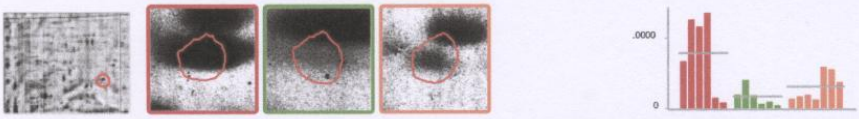
ID	141
Anova	2.848e-6
Presence	100 %
Fold	5.07
Volume	1206



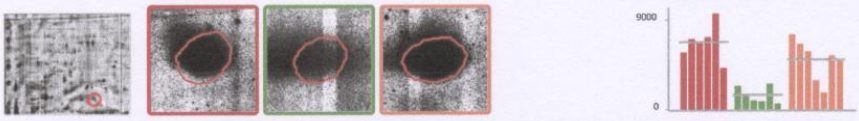
ID	300
Anova	0.0233
Presence	100 %
Fold	4.99
Volume	1999



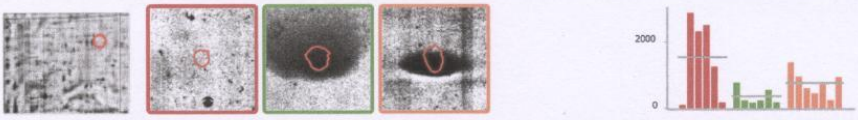
ID	7
Anova	0.0176
Presence	100 %
Fold	4.69
Volume	4147



ID	104
Anova	1.734e-4
Presence	100 %
Fold	4.62
Volume	4452



ID	3
Anova	0.0365
Presence	100 %
Fold	4.32
Volume	879



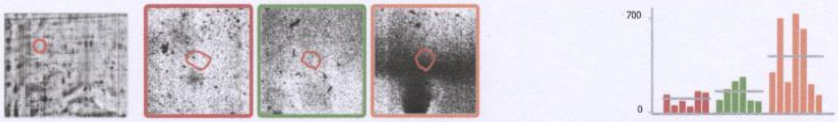
ID	253
Anova	0.0138
Presence	100 %
Fold	4.30
Volume	541

Item 1 to 10 of 61

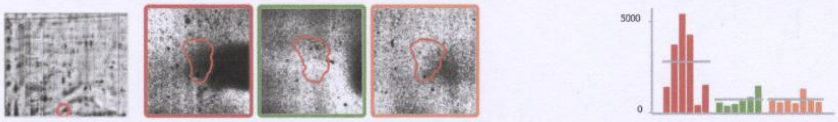
LUDESI

04/01/2010

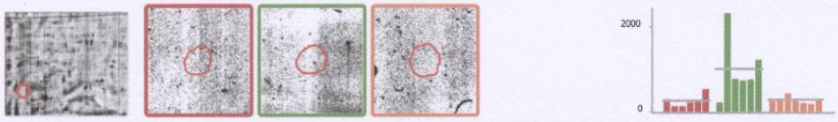
ID	893
Anova	0.0074
Presence	100 %
Fold	4.00
Volume	237



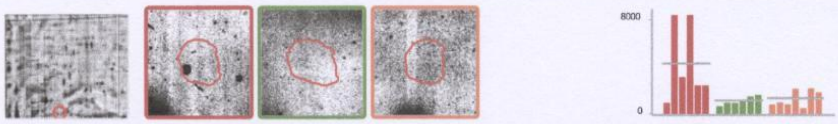
ID	558
Anova	0.0072
Presence	100 %
Fold	3.88
Volume	1378



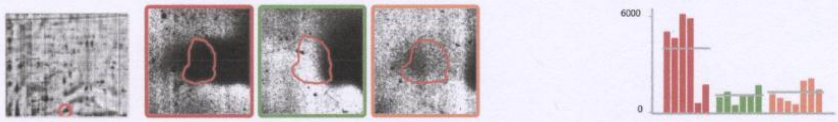
ID	709
Anova	0.0091
Presence	100 %
Fold	3.85
Volume	508



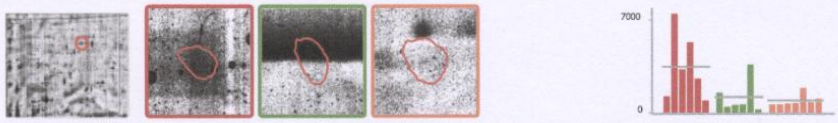
ID	126
Anova	0.0162
Presence	100 %
Fold	3.78
Volume	2211



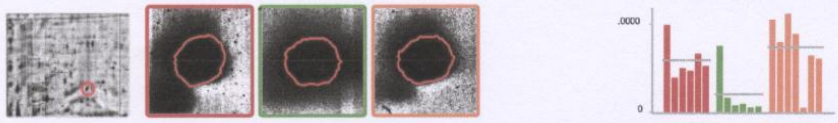
ID	267
Anova	0.0025
Presence	100 %
Fold	3.76
Volume	2064



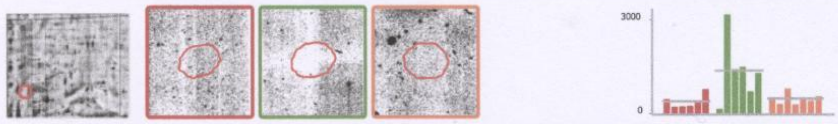
ID	24
Anova	0.0242
Presence	100 %
Fold	3.74
Volume	1818



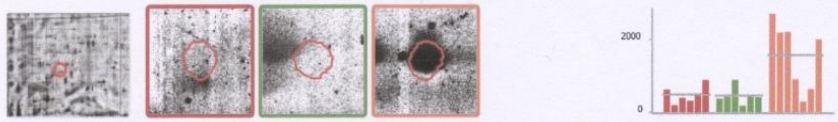
ID	149
Anova	0.0139
Presence	100 %
Fold	3.66
Volume	5225



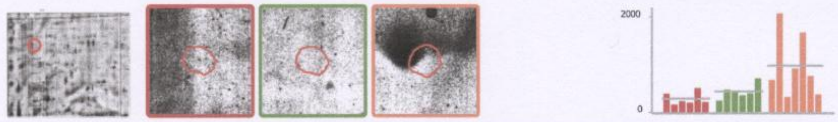
ID	645
Anova	0.0185
Presence	100 %
Fold	3.55
Volume	731



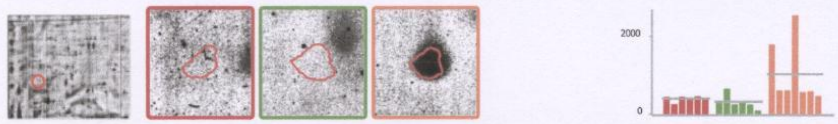
ID	415
Anova	0.0055
Presence	100 %
Fold	3.43
Volume	874



ID	402
Anova	0.0220
Presence	100 %
Fold	3.43
Volume	587



ID	776
Anova	0.0422
Presence	100 %
Fold	3.26
Volume	605

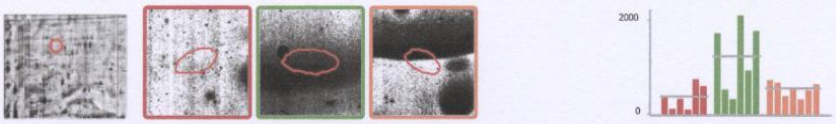


Item 11 to 21 of 61

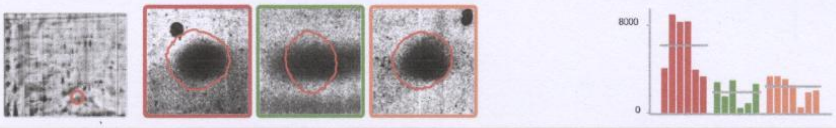
LUDESI

04/01/2010

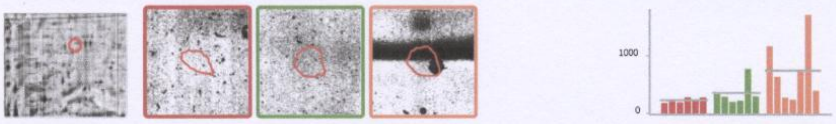
ID	578
Anova	0.0101
Presence	100 %
Fold	3.26
Volume	711



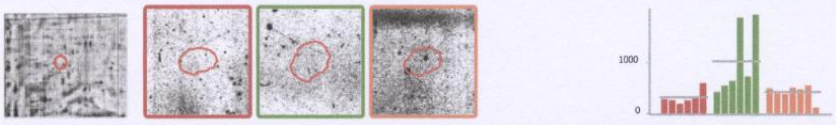
ID	80
Anova	8.341e-4
Presence	100 %
Fold	3.26
Volume	3433



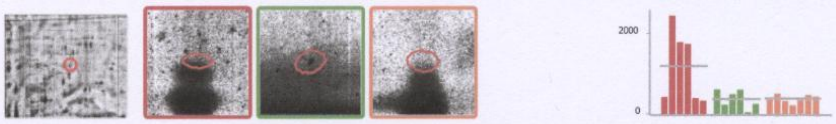
ID	983
Anova	0.0431
Presence	100 %
Fold	3.24
Volume	460



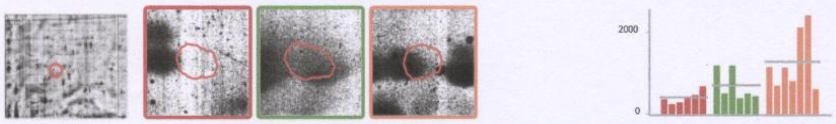
ID	573
Anova	0.0156
Presence	100 %
Fold	3.24
Volume	581



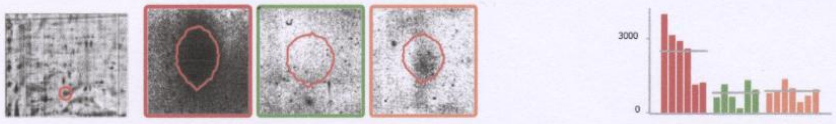
ID	342
Anova	0.0239
Presence	100 %
Fold	3.23
Volume	629



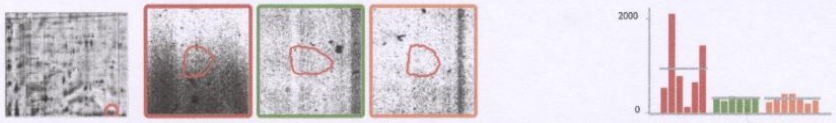
ID	512
Anova	0.0171
Presence	100 %
Fold	3.11
Volume	829



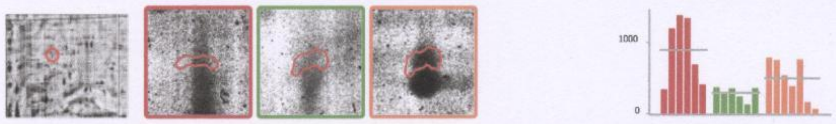
ID	547
Anova	7.293e-4
Presence	100 %
Fold	3.11
Volume	1365



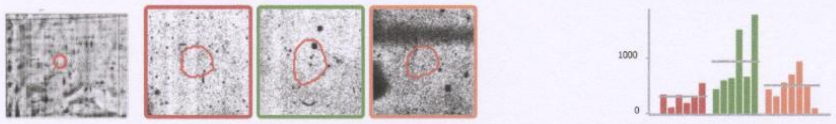
ID	276
Anova	0.0205
Presence	100 %
Fold	3.07
Volume	507



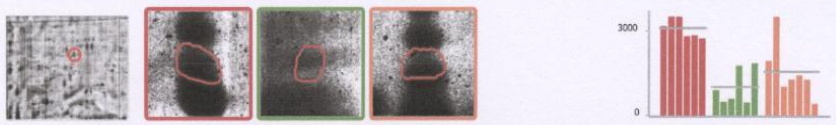
ID	1074
Anova	0.0163
Presence	100 %
Fold	3.07
Volume	557



ID	603
Anova	0.0239
Presence	100 %
Fold	3.07
Volume	581



ID	399
Anova	3.558e-4
Presence	100 %
Fold	3.07
Volume	1869

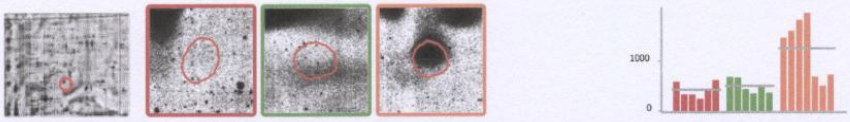


Item 22 to 32 of 61

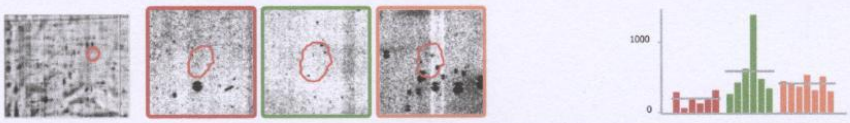
LUDESI

04/01/2010

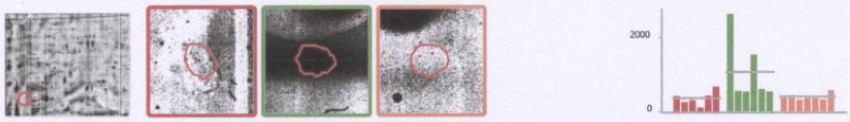
ID	768
Anova	0.0021
Presence	100 %
Fold	2.96
Volume	754



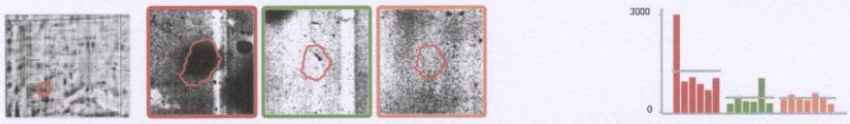
ID	610
Anova	0.0449
Presence	100 %
Fold	2.95
Volume	399



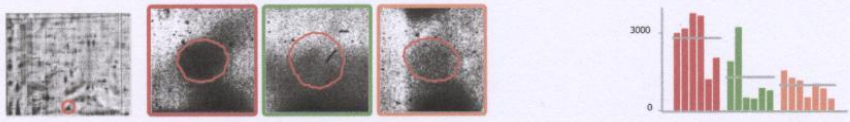
ID	991
Anova	0.0413
Presence	100 %
Fold	2.95
Volume	600



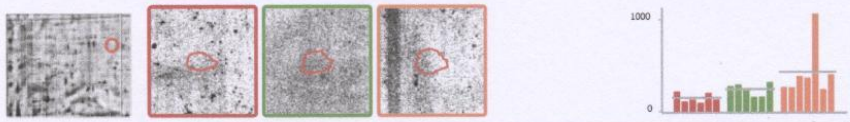
ID	960
Anova	0.0165
Presence	100 %
Fold	2.92
Volume	706



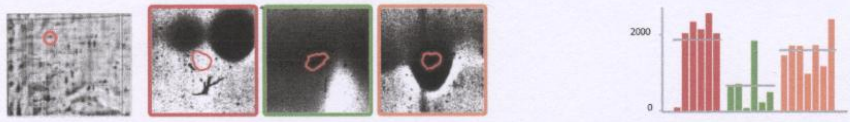
ID	270
Anova	0.0037
Presence	100 %
Fold	2.89
Volume	1649



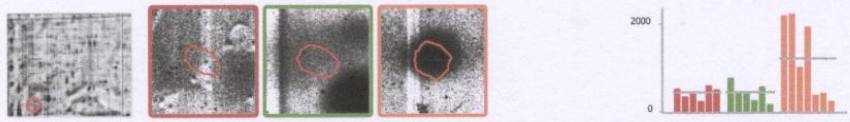
ID	753
Anova	0.0383
Presence	100 %
Fold	2.88
Volume	283



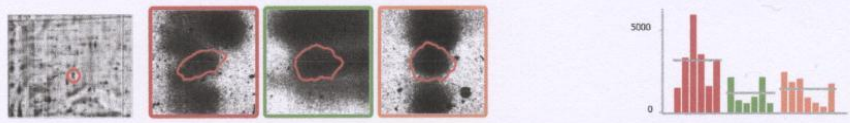
ID	6
Anova	0.0164
Presence	100 %
Fold	2.84
Volume	1385



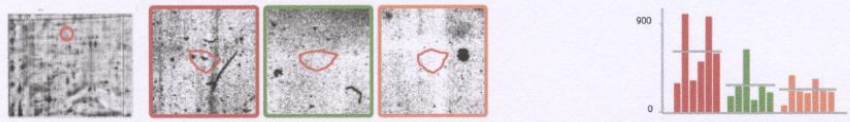
ID	1008
Anova	0.0378
Presence	100 %
Fold	2.75
Volume	727



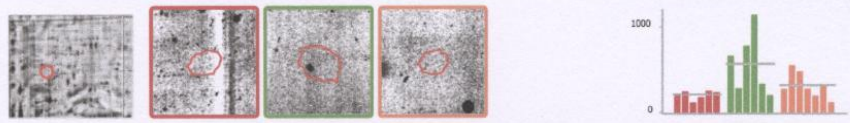
ID	218
Anova	0.0126
Presence	100 %
Fold	2.71
Volume	1895



ID	524
Anova	0.0096
Presence	100 %
Fold	2.70
Volume	364



ID	1004
Anova	0.0381
Presence	100 %
Fold	2.70
Volume	364

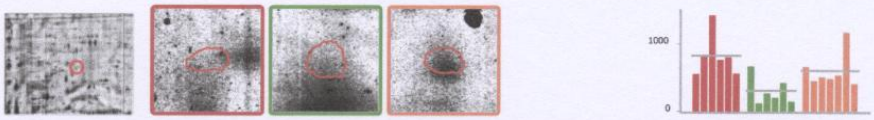


Item 33 to 43 of 61

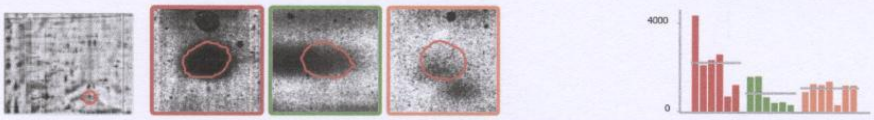
LUDESI

04/01/2010

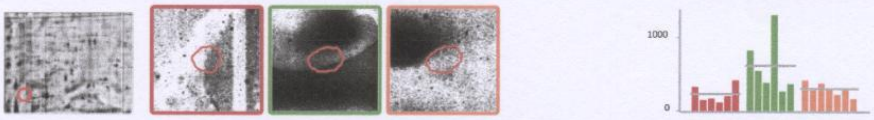
ID	815
Anova	0.0136
Presence	100 %
Fold	2.69
Volume	580



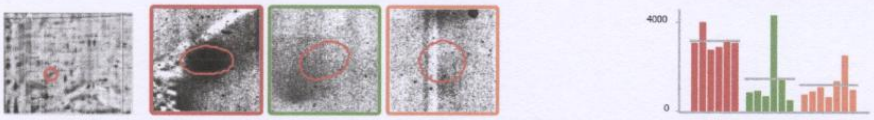
ID	305
Anova	0.0211
Presence	100 %
Fold	2.68
Volume	1346



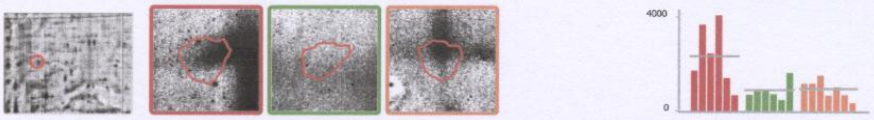
ID	891
Anova	0.0252
Presence	100 %
Fold	2.67
Volume	379



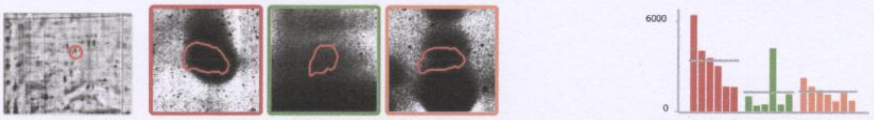
ID	244
Anova	0.0034
Presence	100 %
Fold	2.66
Volume	1905



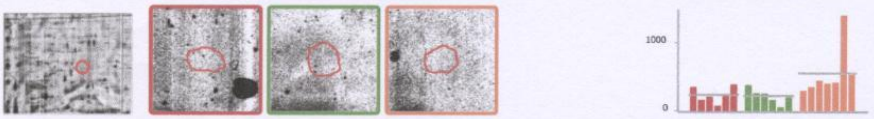
ID	287
Anova	0.0101
Presence	100 %
Fold	2.65
Volume	1357



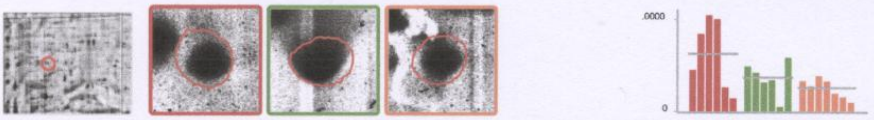
ID	708
Anova	0.0211
Presence	100 %
Fold	2.62
Volume	1976



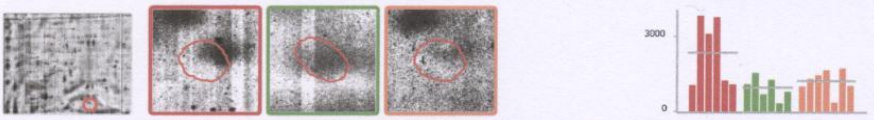
ID	819
Anova	0.0475
Presence	100 %
Fold	2.50
Volume	348



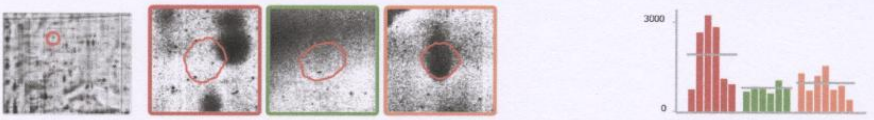
ID	34
Anova	0.0479
Presence	100 %
Fold	2.47
Volume	4074



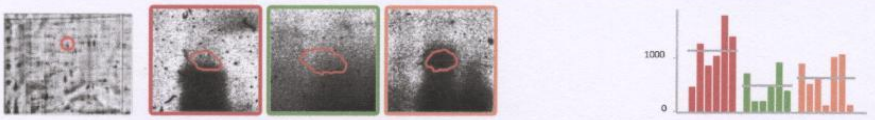
ID	238
Anova	0.0271
Presence	100 %
Fold	2.46
Volume	1498



ID	140
Anova	0.0226
Presence	100 %
Fold	2.41
Volume	1211



ID	1063
Anova	0.0250
Presence	100 %
Fold	2.34
Volume	746

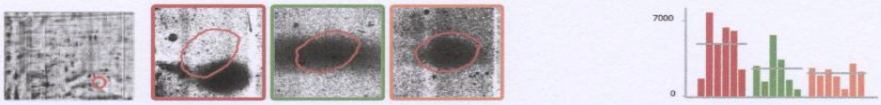


Item 44 to 54 of 61

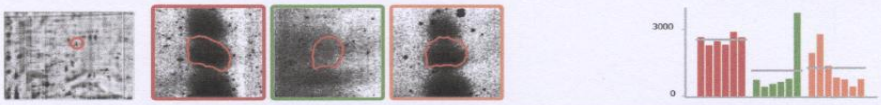
LUDESI

04/01/2010

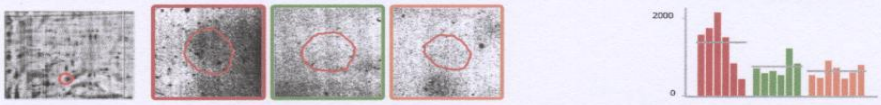
ID	87
Anova	0.0301
Presence	100 %
Fold	2.28
Volume	3128



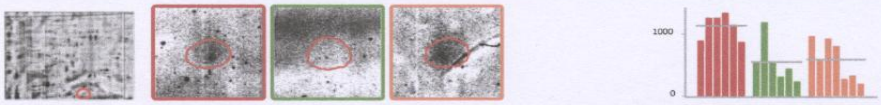
ID	466
Anova	0.0264
Presence	100 %
Fold	2.22
Volume	1636



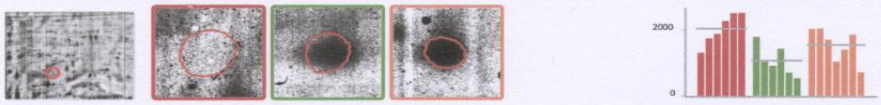
ID	539
Anova	0.0096
Presence	100 %
Fold	2.14
Volume	907



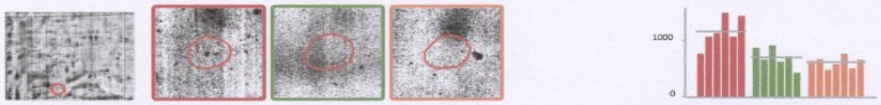
ID	712
Anova	0.0052
Presence	100 %
Fold	2.06
Volume	748



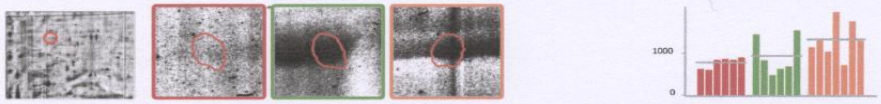
ID	563
Anova	0.0120
Presence	100 %
Fold	1.91
Volume	1553



ID	531
Anova	2.280e-4
Presence	100 %
Fold	1.90
Volume	810



ID	877
Anova	0.0441
Presence	100 %
Fold	1.70
Volume	1036



Item 55 to 61 of 61

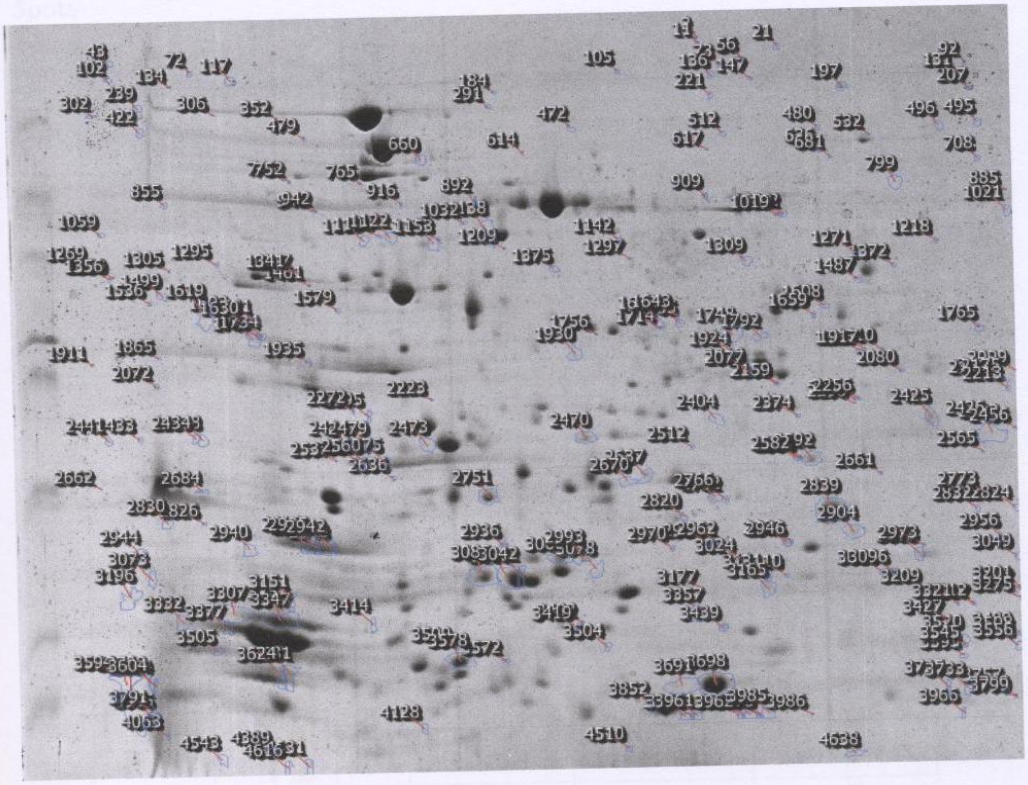
F3 Gel analysis data for *E. cancerogenus* and *B. fragilis* using NonLinear Dynamics SameSpots software

PROTBF 1

Experiment: PROTBF 1

Report created: 16/01/2010 22:18:43

Reference image



PROTBF 1

Experiment: PROTBF 1

Report created: 20/01/2010 23:53:23

Spots

#	Anova (p)	Fold	Tags	Notes	Average Normalised Volumes		
					Non mucin media	Mucin Type II media	Mucin Type III media
3698	3.043e-004	6.0			3.148e+006	7.924e+005	5.241e+005
2484	0.002	3.2			1.516e+005	4.788e+004	6.999e+004
2470	0.003	2.2			2.941e+005	2.068e+005	1.350e+005
3006	0.006	2.3			2.202e+005	5.007e+005	2.233e+005
1608	0.011	2.7			1.558e+004	2.609e+004	4.131e+004
2592	0.017	2.1			3.104e+005	1.456e+005	1.927e+005
1138	0.024	3.4			2.021e+005	6.919e+005	4.931e+005
2584	0.025	1.6			4950.707	3996.665	3040.097
3691	0.028	2.8			6.981e+005	9.814e+005	3.530e+005
1295	0.029	6.5			3192.412	2.091e+004	5940.371
3572	0.030	1.8			1.101e+004	1.267e+004	6902.432
4128	0.031	3.0			3.009e+004	9.073e+004	6.460e+004
4319	0.033	6.0			8.717e+004	3.030e+005	5.204e+005
3986	0.054	7.7			1.361.911	1.043e+004	2036.838

Tags	
	Power spots
	p value spots
	fold change of 3.0 or more
	p value spots

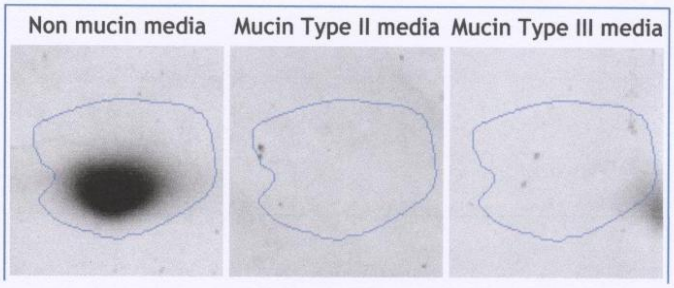
Identifier 3698

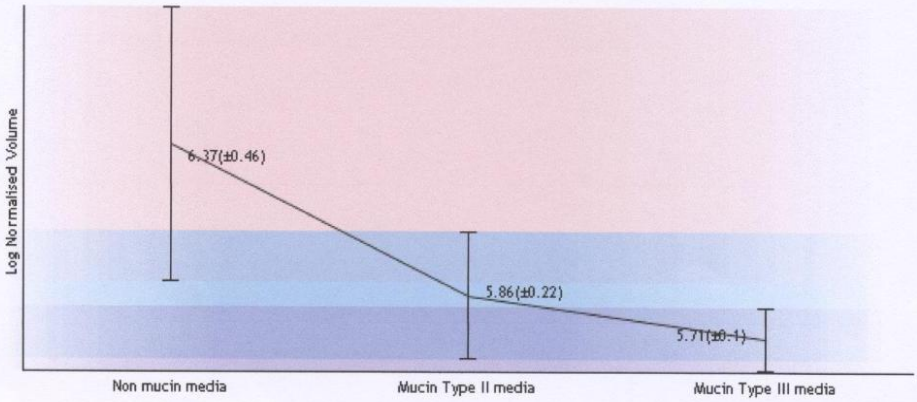
Position (2138, 2189)

Notes

- fold change of 3.0 or more
- p value spots
- p value spots

Power spots



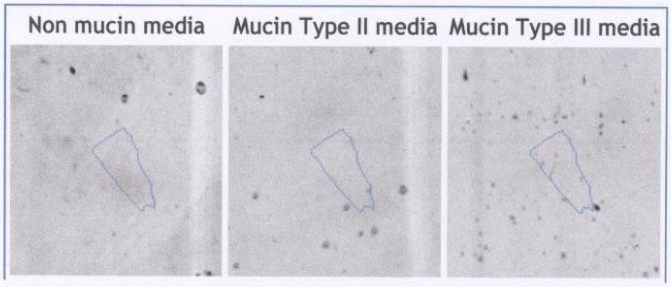


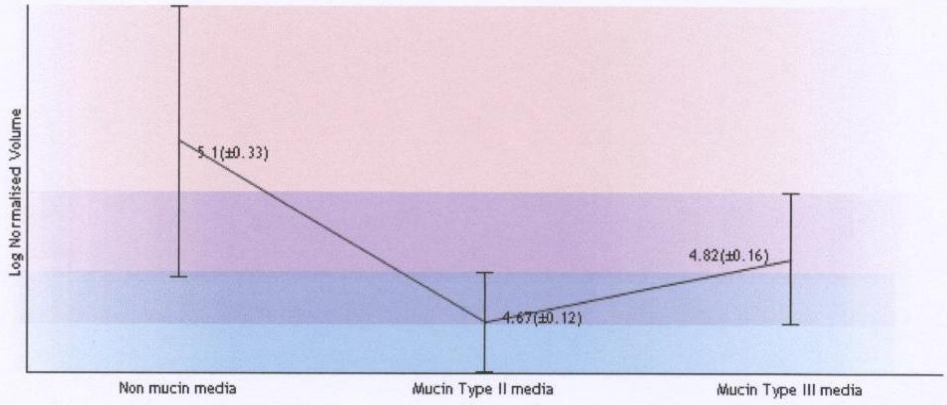
Identifier 2484

Position (1039, 1385)

Notes

- fold change of 3.0 or more
- p value spots
- p value spots
- Power spots



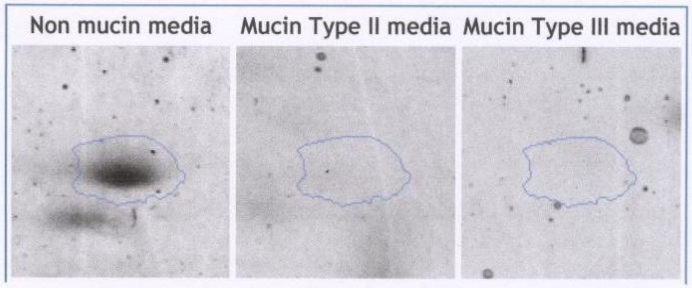


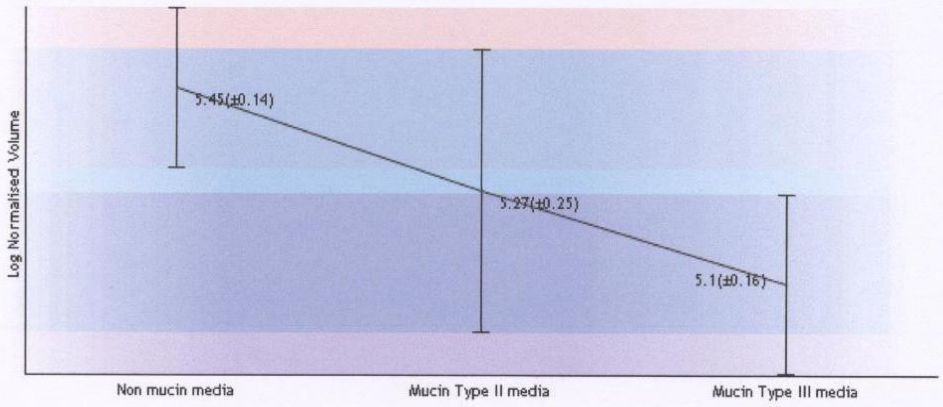
Identifier 2470

Position (1760, 1373)

Notes

- p value spots
- p value spots
- Power spots



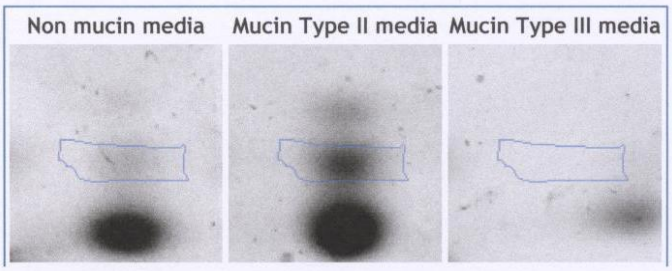


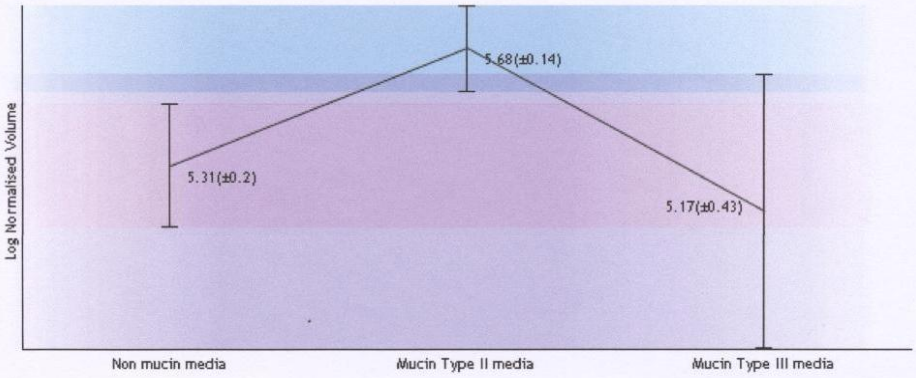
Identifier 3006

Position (1674, 1761)

Notes

- p value spots
- p value spots
- Power spots



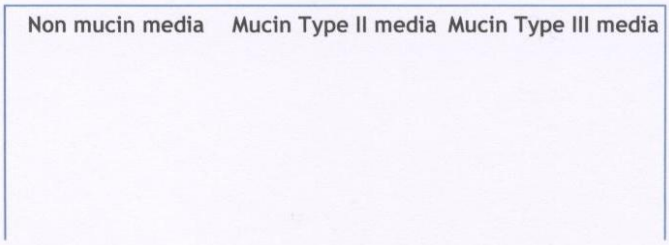


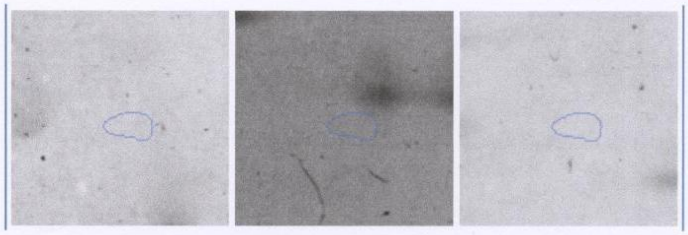
Identifier 1608

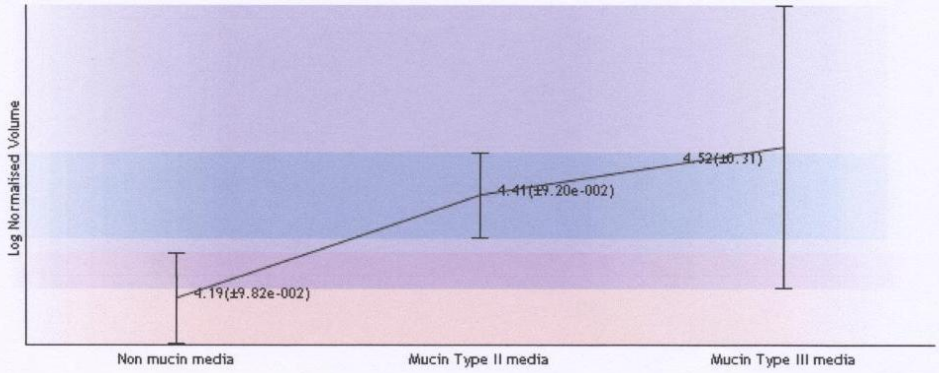
Position (2509, 958)

Notes

- p value spots
- p value spots
- Power spots





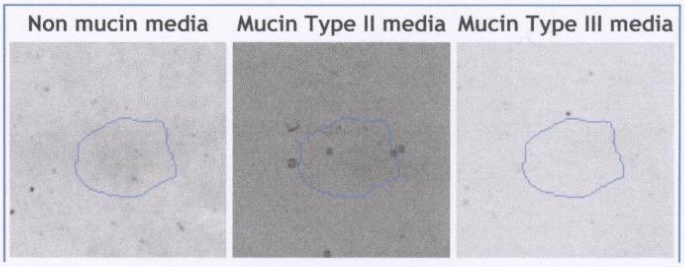


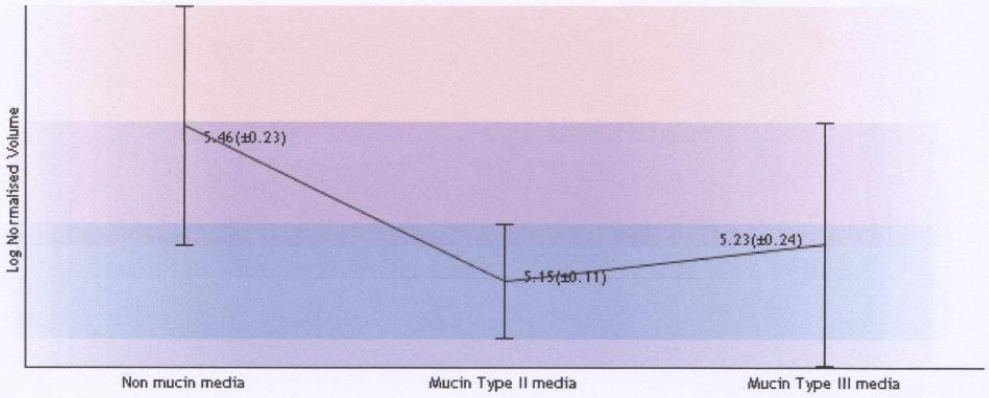
Identifier 2592

Position (2452, 1455)

Notes

- p value spots
- p value spots



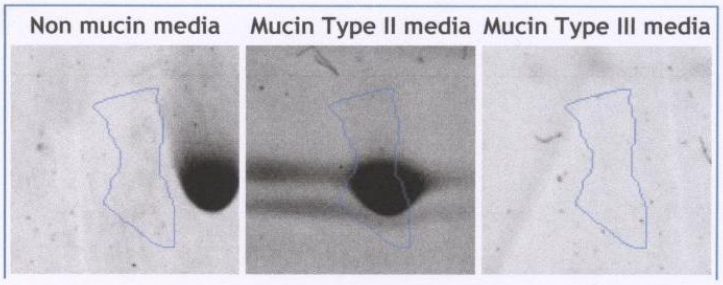


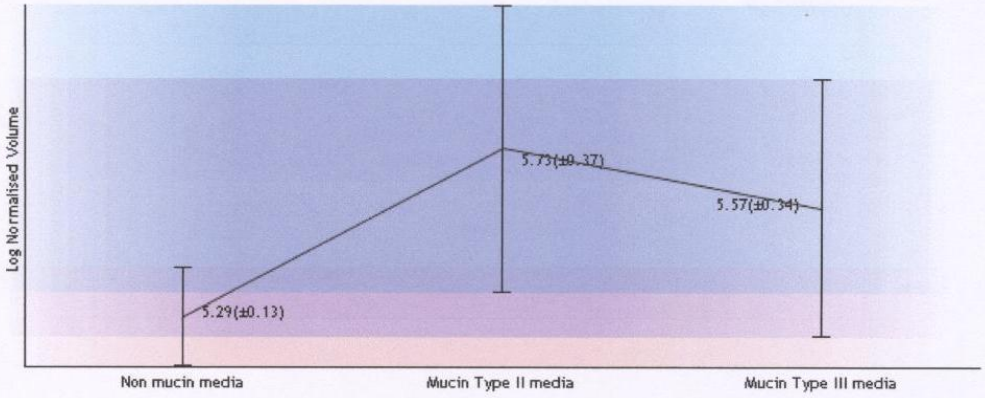
Identifier 1138

Position (1489, 721)

Notes

- fold change of 3.0 or more
- p value spots
- p value spots



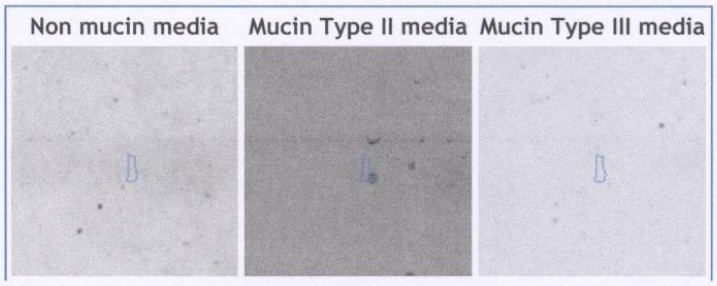


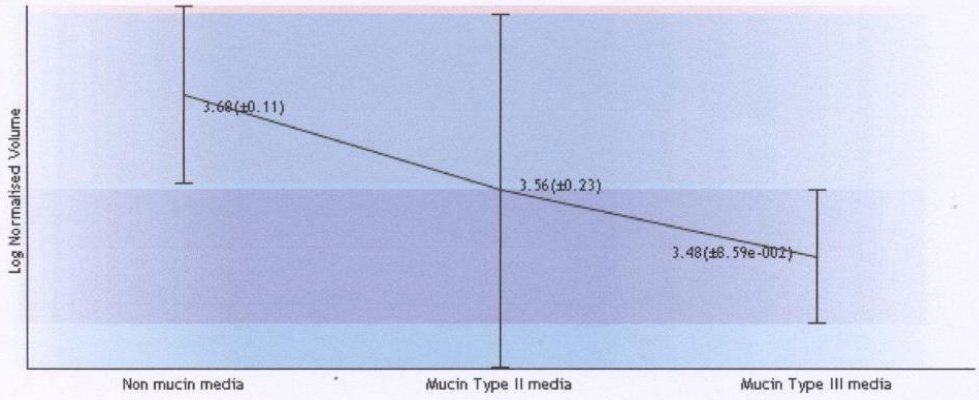
Identifier 2584

Position (2415, 1450)

Notes

- p value spots
- p value spots



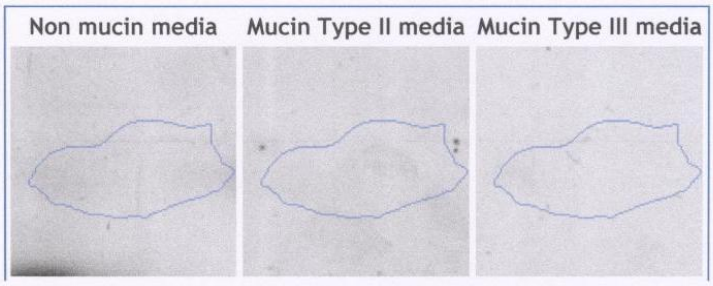


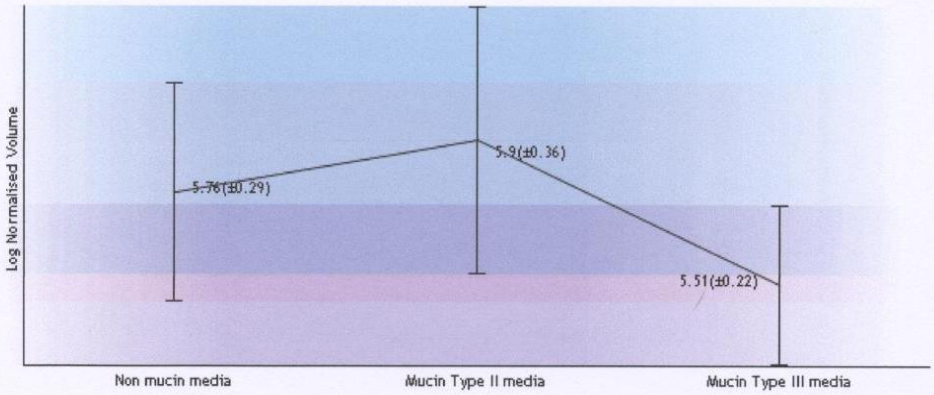
Identifier 3691

Position (2047, 2183)

Notes

- p value spots
- p value spots



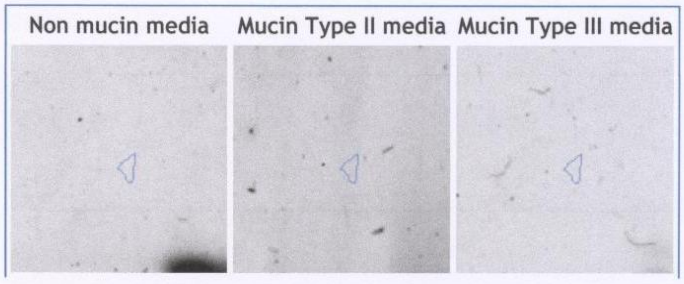


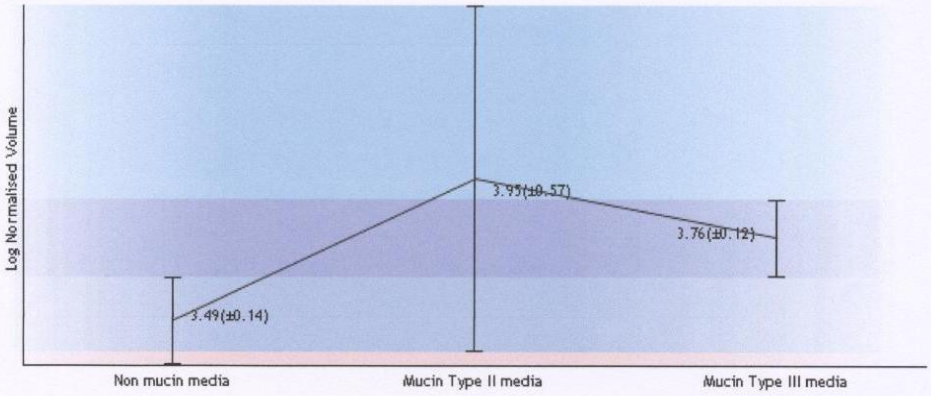
Identifier 1295

Position (639, 795)

Notes

- fold change of 3.0 or more
- p value spots
- p value spots



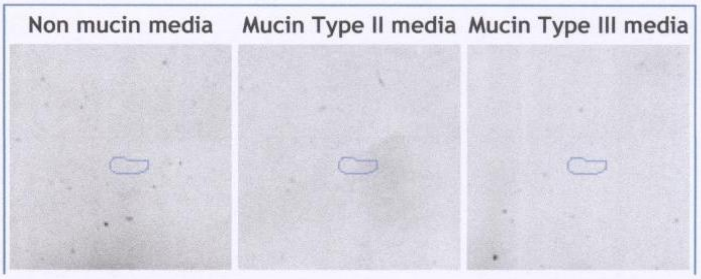


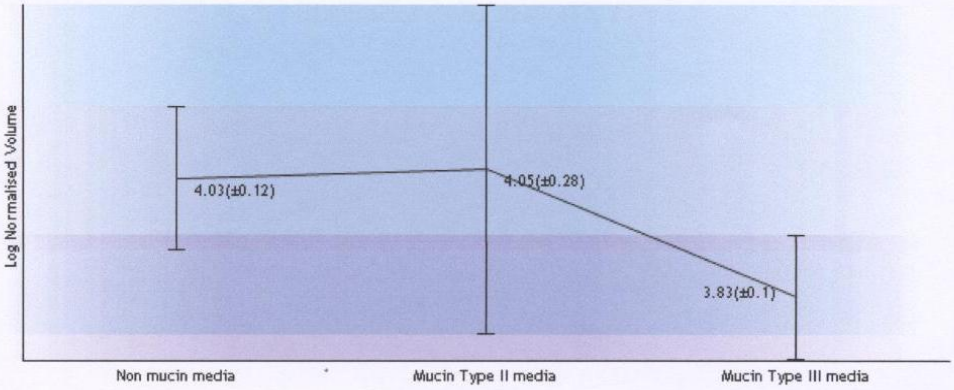
Identifier 3572

Position (1501, 2083)

Notes

- p value spots
- p value spots



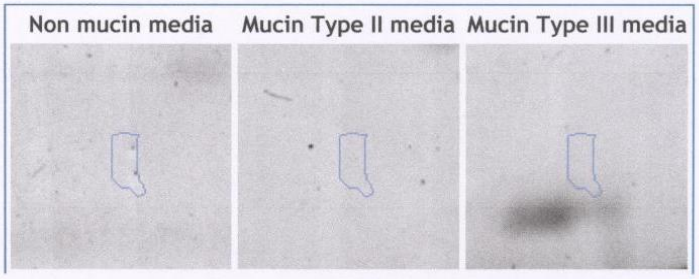


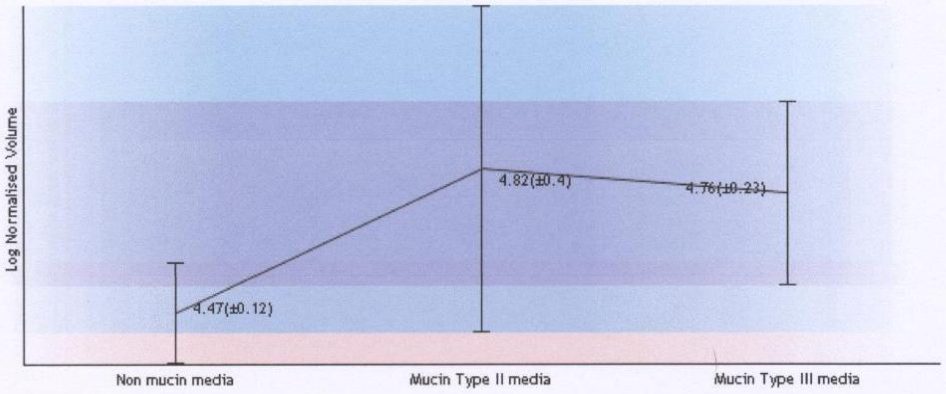
Identifier 4128

Position (1241, 2305)

Notes

- fold change of 3.0 or more
- p value spots
- p value spots



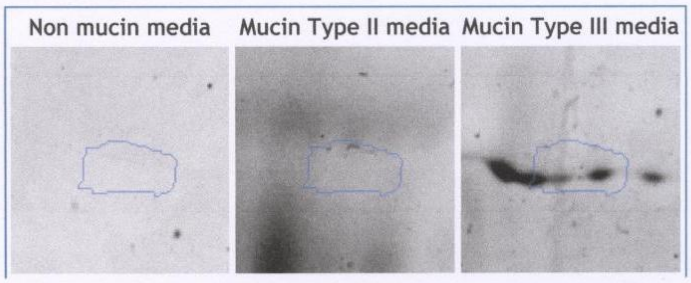


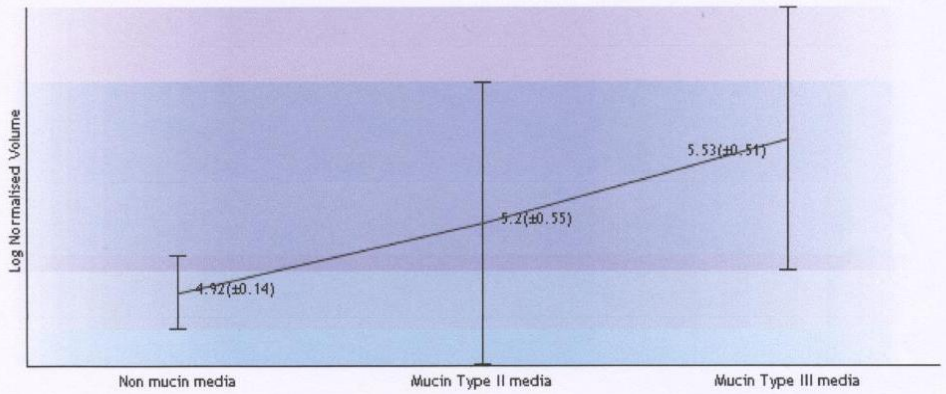
Identifier 4319

Position (565, 2343)

Notes

- fold change of 3.0 or more
- p value spots
- p value spots



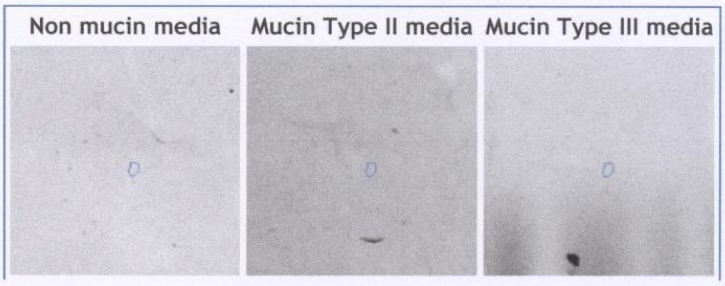


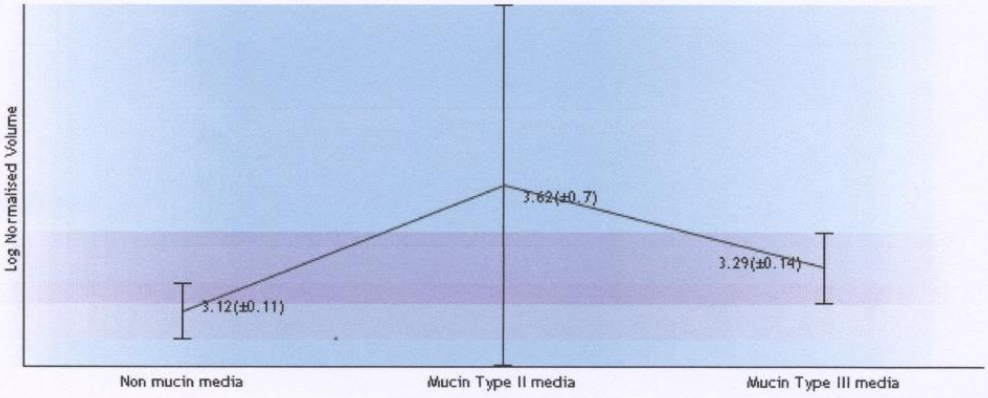
Identifier 3986

Position (2438, 2281)

Notes

- fold change of 3.0 or more
- p value spots
- p value spots



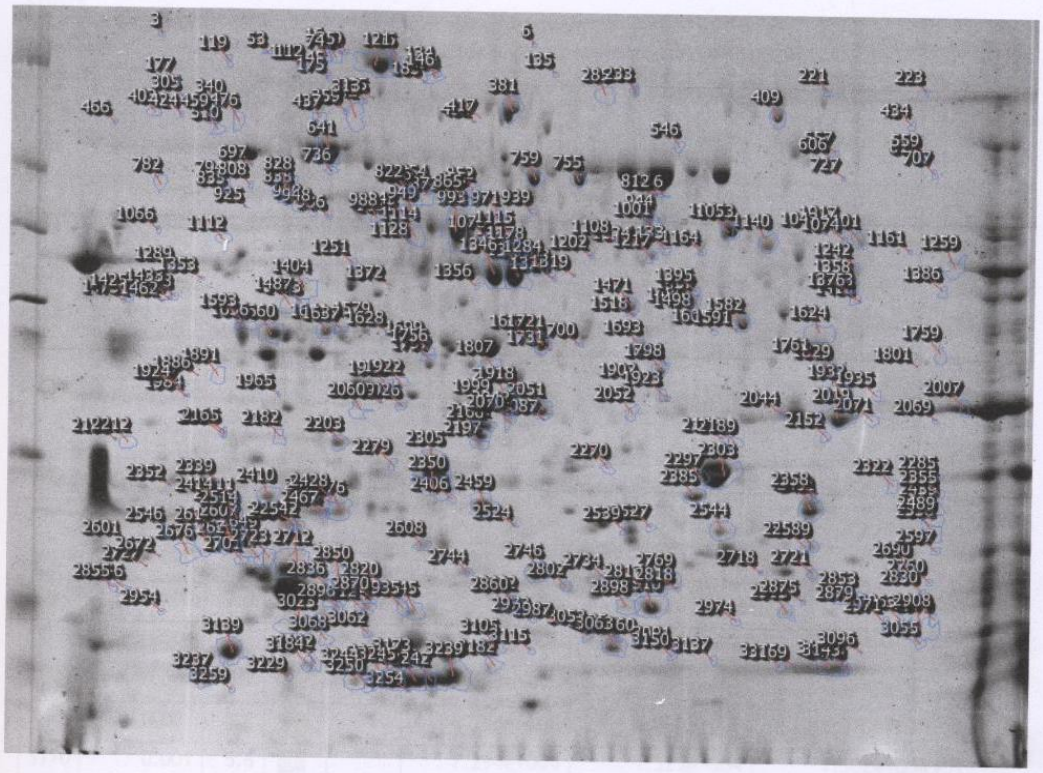


PROTEOMICS 3

Experiment: PROTEOMICS 3

Report created: 16/01/2010 22:22:41

Reference image



PROTEOMICS 3

Experiment: PROTEOMICS 3

Report created: 20/01/2010 23:57:30

Spots

#	Anova (p)	Fold	Tags	Notes	Average Normalised Volumes		
					Non mucin media	Mucin type II media	Mucin type III media
95	1.039e-004	3.0			3.023e+005	1.031e+005	1.012e+005
2914	1.114e-004	2.3			1.123e+005	4.946e+004	5.120e+004
2026	1.362e-004	1.9			6.443e+004	3.867e+004	3.409e+004
2132	3.050e-004	4.8			4.420e+005	1.284e+005	9.248e+004
755	3.761e-004	6.7			5.050e+005	7.543e+004	8.154e+004
3176	5.017e-004	6.1			1.166e+006	2.160e+005	1.925e+005
2484	6.287e-004	2.0			2.108e+004	1.329e+004	1.052e+004
2717	6.672e-004	5.1			6.585e+004	3.377e+005	8.572e+004
2945	6.762e-004	2.3			2.858e+005	1.318e+005	1.254e+005
2365	8.818e-004	2.9			2.652e+004	9878.909	9019.675
2608	0.001	2.7			1.663e+005	6.131e+004	7.626e+004
2746	0.001	3.8			9.243e+005	2.420e+005	2.470e+005
1171	0.002	3.0			3.900e+005	1.467e+005	1.303e+005
3139	0.002	3.3			1.783e+006	5.443e+005	5.895e+005
854	0.002	3.1			9.330e+004	2.854e+005	2.040e+005
1234	0.002	3.2			4.140e+005	1.294e+005	1.517e+005
3234	0.002	2.5			2.018e+005	8.117e+004	1.190e+005
2352	0.003	1.9			3754.338	1944.626	3165.534
2410	0.003	3.7			5.802e+005	5.349e+005	1.571e+005
1579	0.004	1.9			1433.976	2319.226	2752.321
1342	0.004	4.9			2.388e+006	4.888e+005	7.282e+005
3060	0.009	1.7			2.947e+004	1.909e+004	3.283e+004
813	0.010	3.0			4591.714	1.368e+004	9587.014
1922	0.022	3.5			4.540e+004	1.605e+005	9.859e+004
872	0.024	4.1			8057.272	2.099e+004	3.320e+004
3244	0.026	1.7			2.928e+005	2.396e+005	1.745e+005
2083	0.028	2.0			8.762e+004	4.398e+004	7.334e+004

1902	0.028	1.7			1753.296	3052.482	2859.892
285	0.037	3.6			9.781e+004	3.564e+005	2.645e+005
1807	0.038	3.2			2.874e+005	2.417e+005	9.005e+004
2712	0.043	2.4			3.925e+006	3.254e+006	1.648e+006
835	0.049	1.5			8994.086	8357.012	6113.549

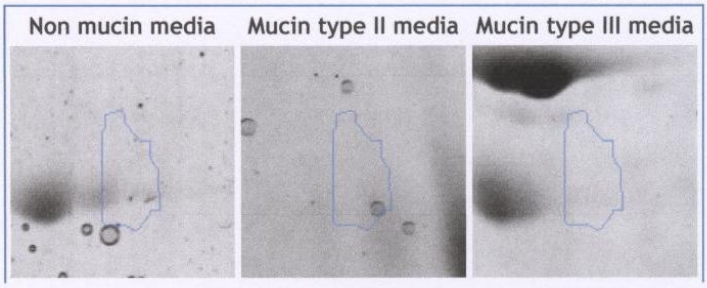
Tags	
	Power spots
	p value spots
	fold change of 3.0 or more
	spots of interest

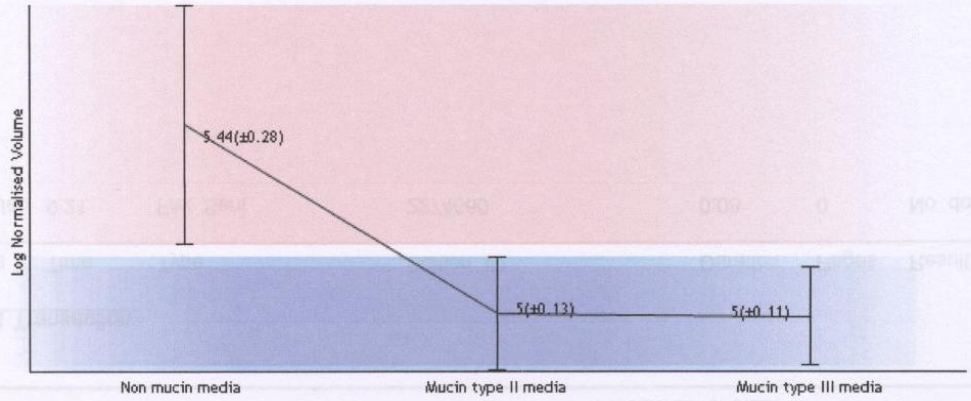
Identifier 95

Position (1149, 163)

Notes

- fold change of 3.0 or more
- p value spots
- Power spots
- spots of interest



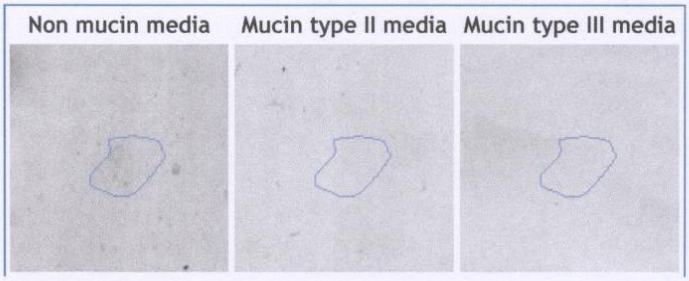


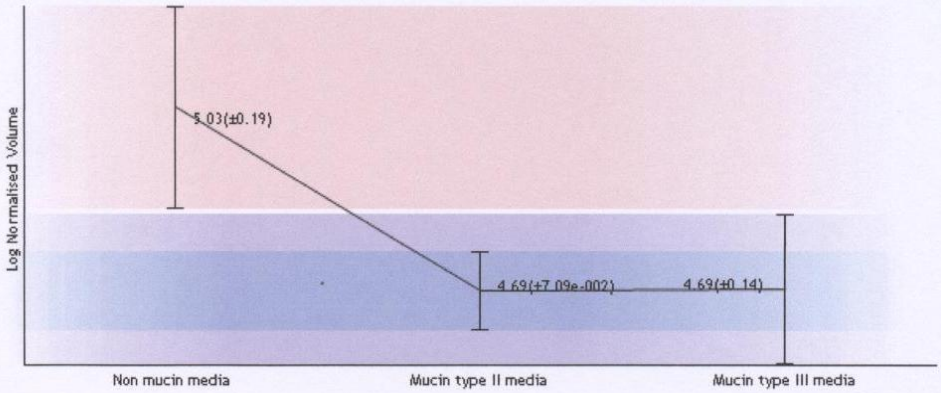
Identifier 2914

Position (1101, 1843)

Notes

- p value spots
- Power spots
- spots of interest



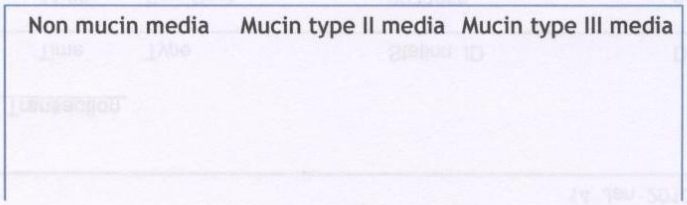


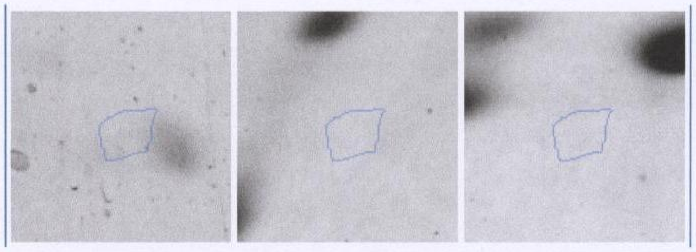
Identifier 2026

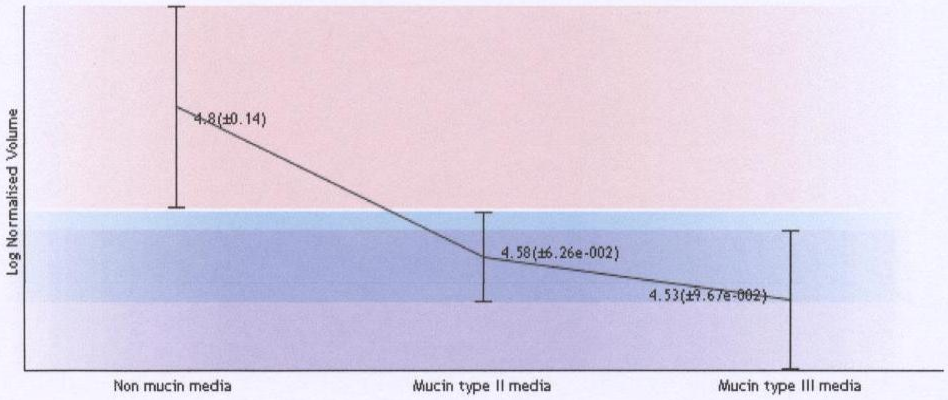
Position (1193, 1219)

Notes

- p value spots
- Power spots
- spots of interest







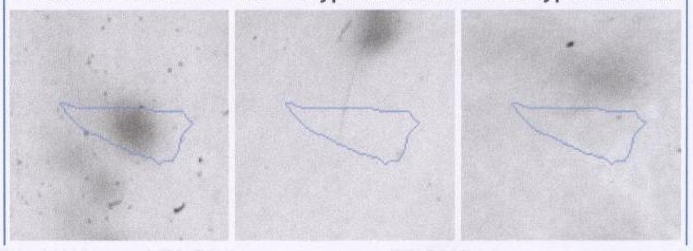
Identifier 2132

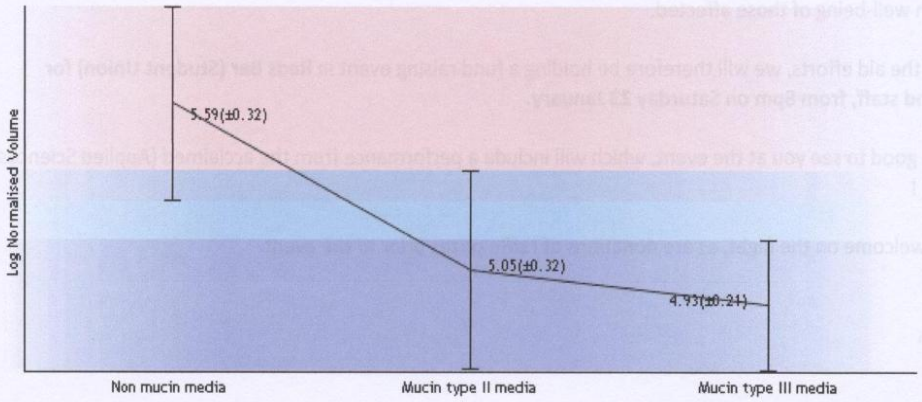
Position (1447, 1279)

Notes

- fold change of 3.0 or more
- p value spots
- Power spots
- spots of interest

Non mucin media	Mucin type II media	Mucin type III media



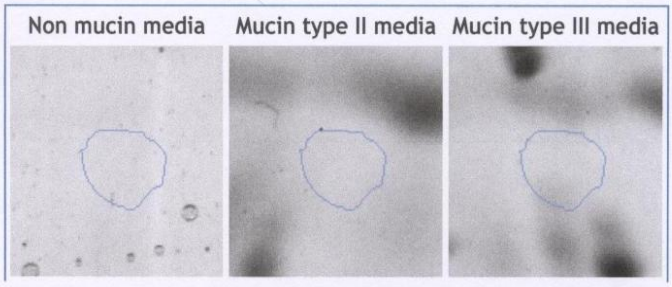


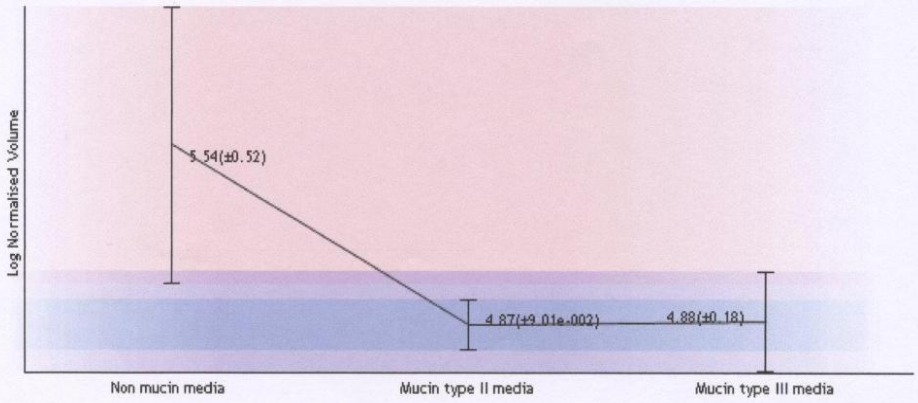
Identifier 755

Position (1721, 521)

Notes

- fold change of 3.0 or more
- p value spots
- Power spots
- spots of interest



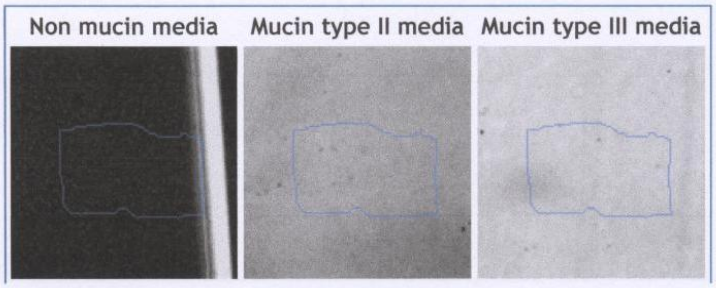


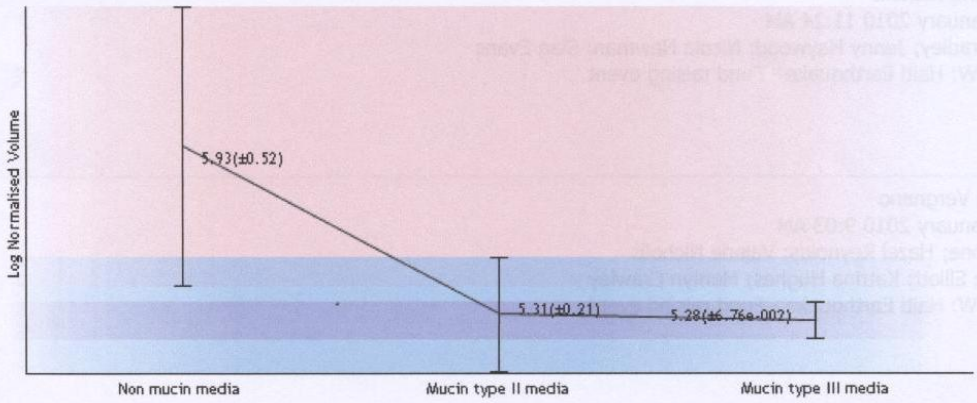
Identifier 3176

Position (2461, 2018)

Notes

- fold change of 3.0 or more
- p value spots
- Power spots
- spots of interest



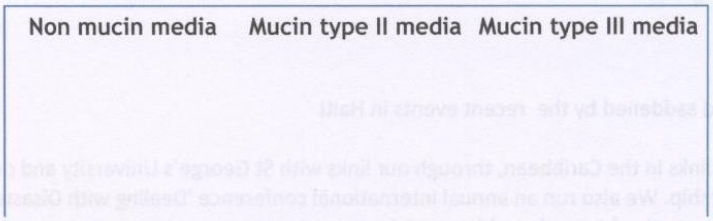


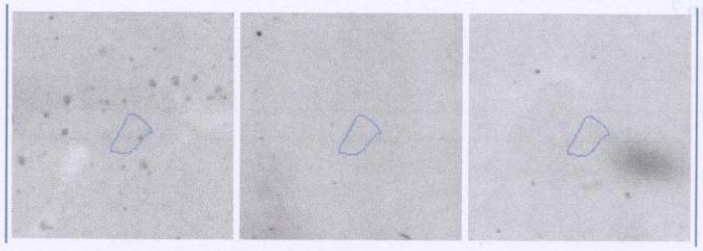
Identifier 2484

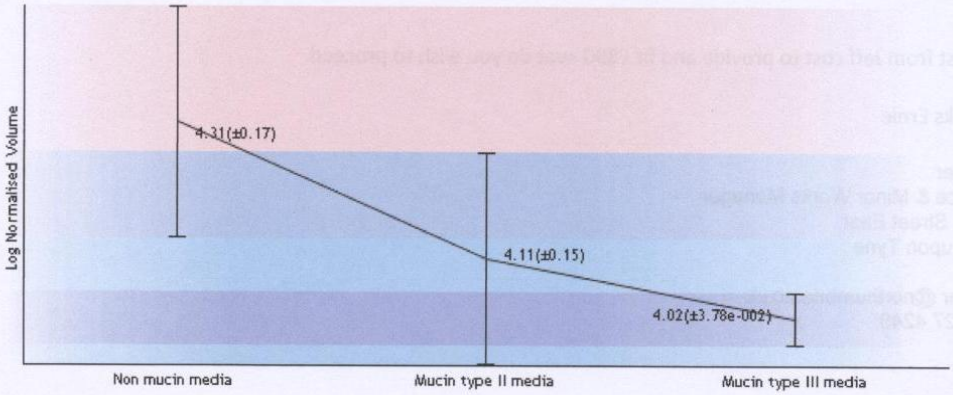
Position (696, 1555)

Notes

- p value spots
- Power spots
- spots of interest





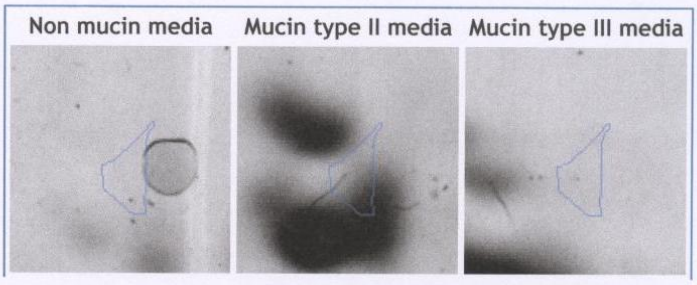


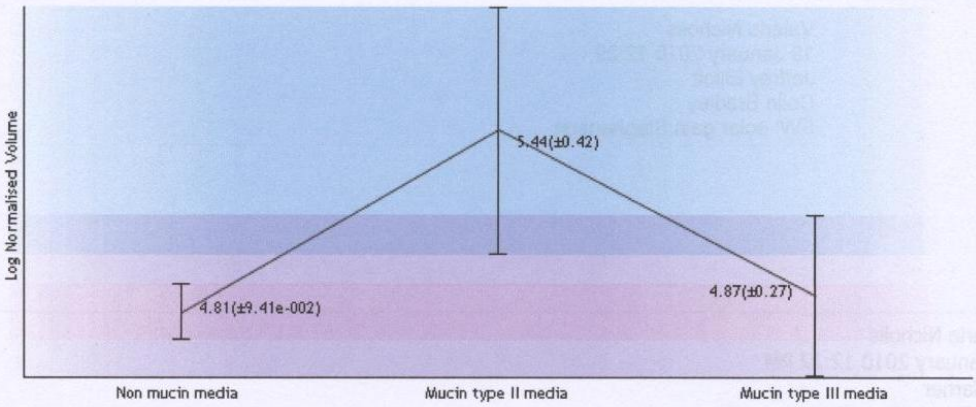
Identifier 2717

Position (778, 1719)

Notes

- fold change of 3.0 or more
- p value spots
- Power spots
- spots of interest



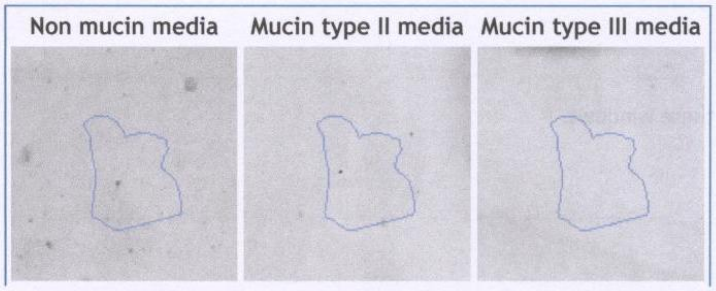


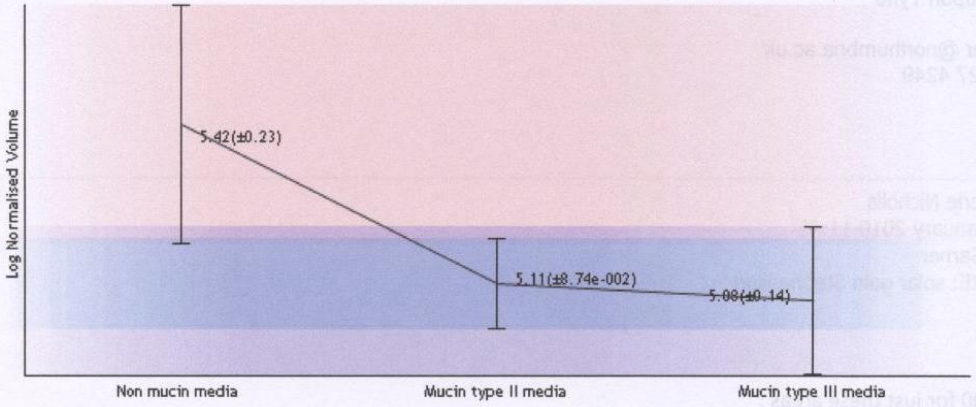
Identifier 2945

Position (1232, 1861)

Notes

- p value spots
- Power spots
- spots of interest



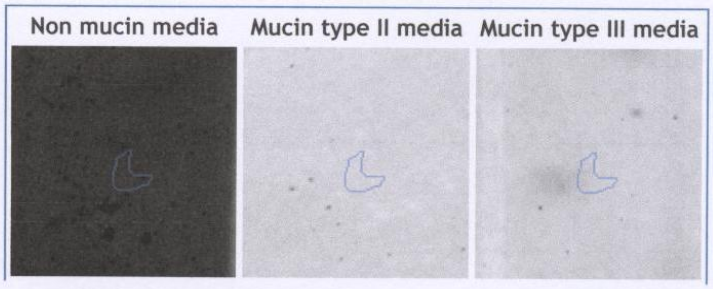


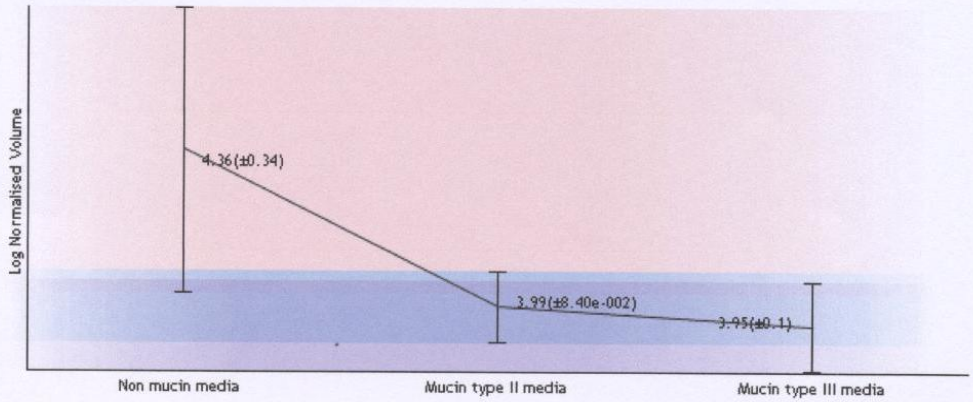
Identifier 2365

Position (2450, 1480)

Notes

- p value spots
- Power spots
- spots of interest



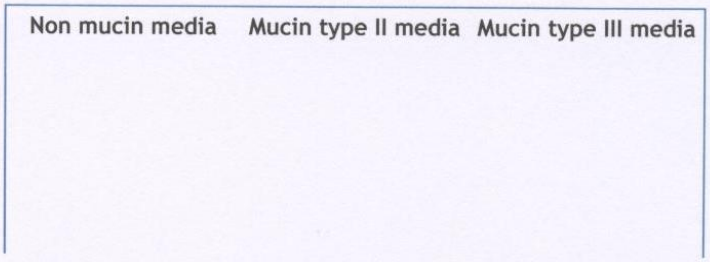


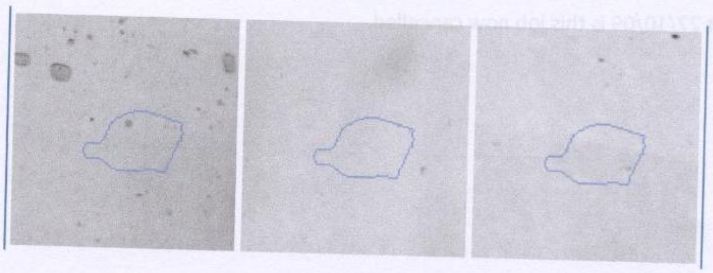
Identifier 2608

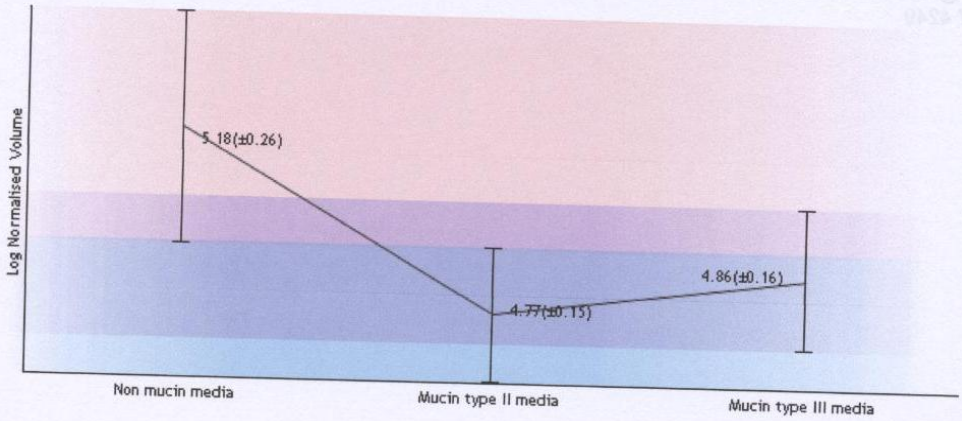
Position (1264, 1650)

Notes

- p value spots
- Power spots
- spots of interest





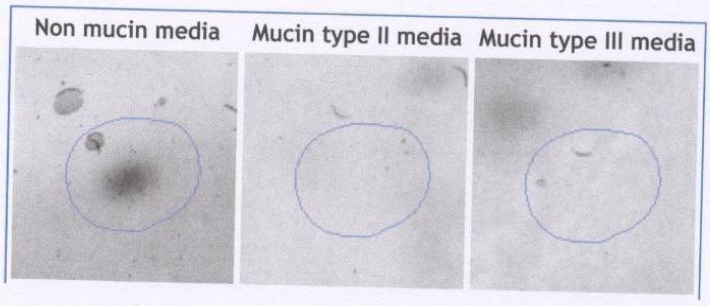


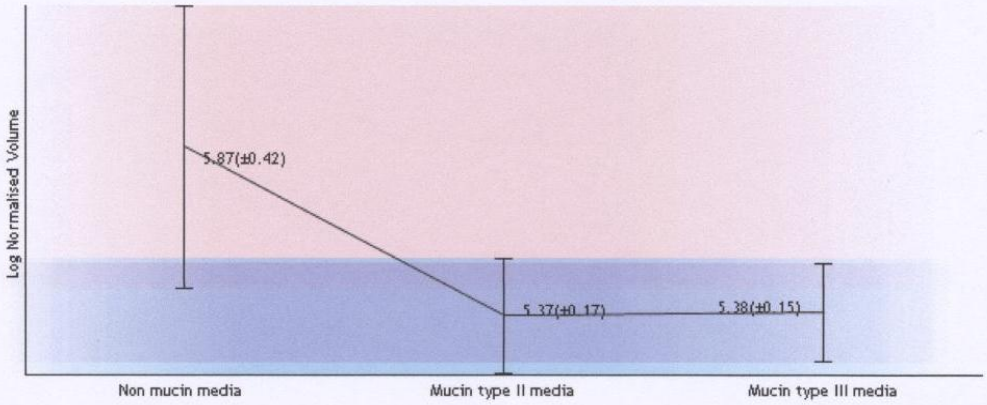
Identifier 2746

Position (1613, 1735)

Notes

- fold change of 3.0 or more
- p value spots
- Power spots
- spots of interest



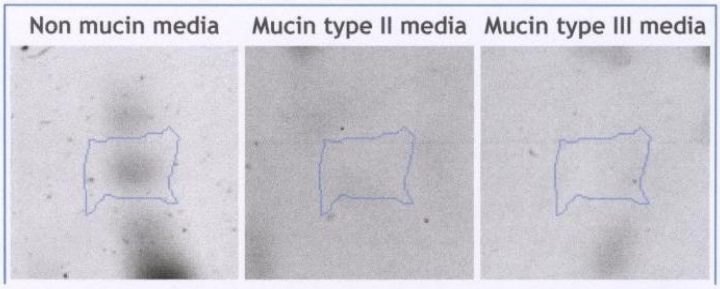


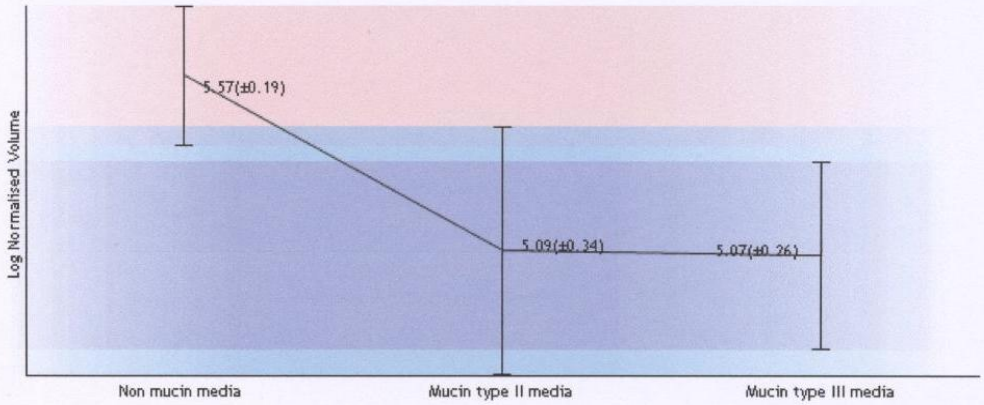
Identifier 1171

Position (1520, 729)

Notes

- fold change of 3.0 or more
- p value spots
- Power spots
- spots of interest



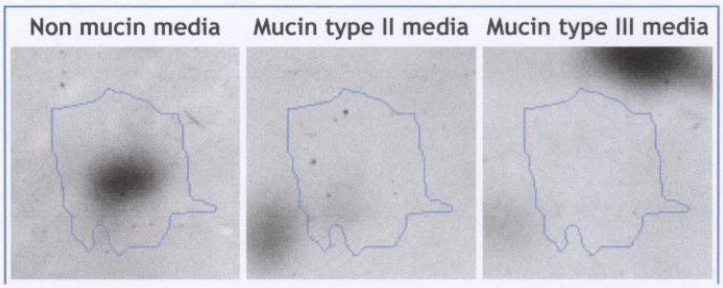


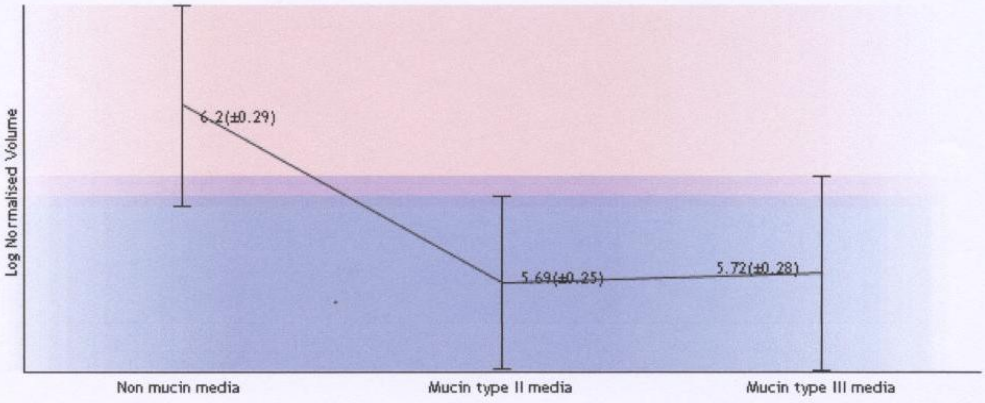
Identifier 3139

Position (685, 1993)

Notes

- fold change of 3.0 or more
- p value spots
- Power spots
- spots of interest



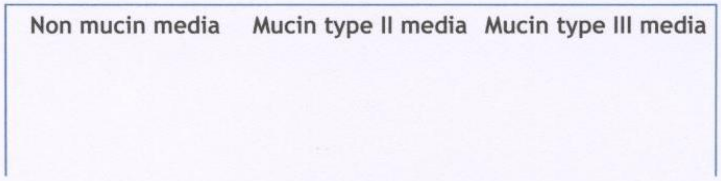


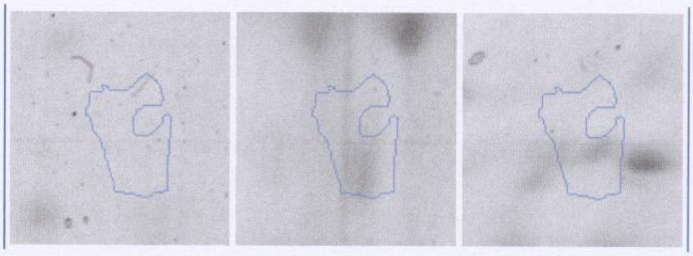
Identifier 854

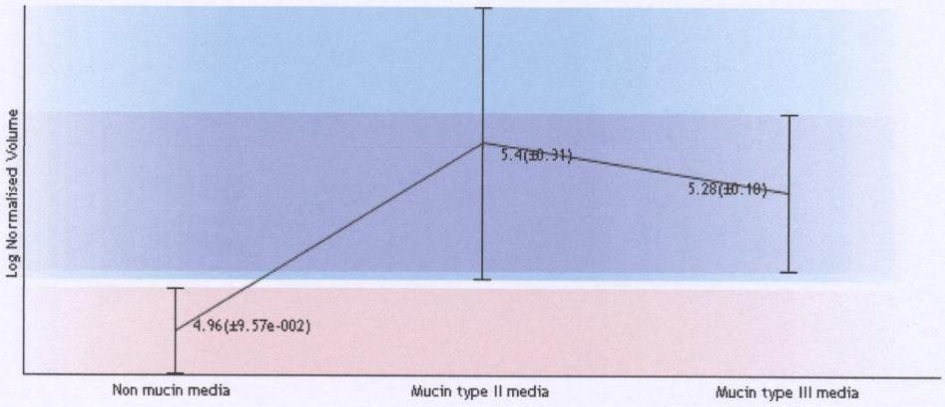
Position (1255, 574)

Notes

- fold change of 3.0 or more
- p value spots
- Power spots
- spots of interest





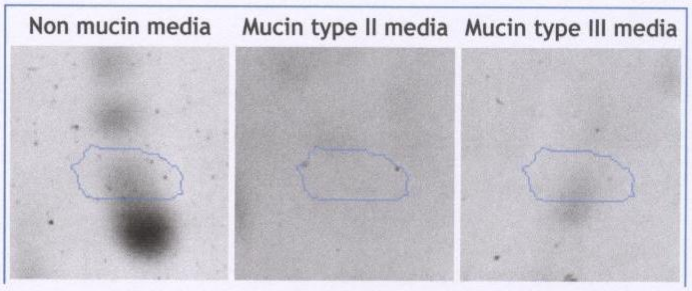


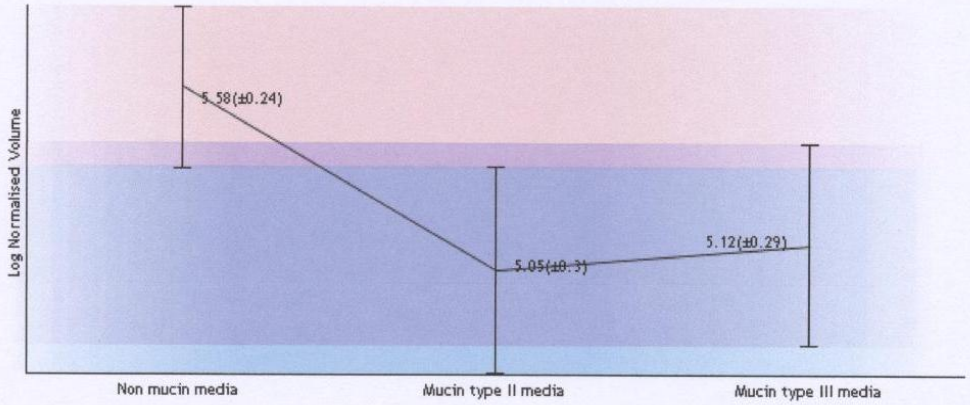
Identifier 1234

Position (1523, 765)

Notes

- fold change of 3.0 or more
- p value spots
- Power spots
- spots of interest



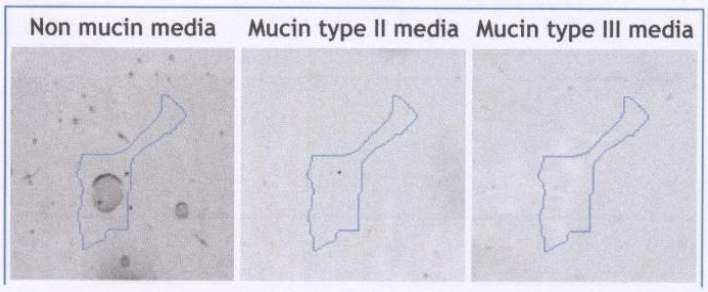


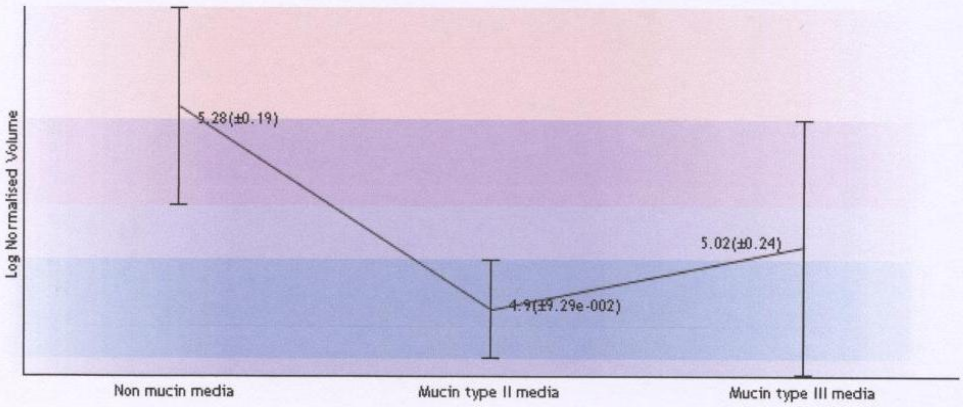
Identifier 3234

Position (1143, 2065)

Notes

- p value spots
- Power spots
- spots of interest



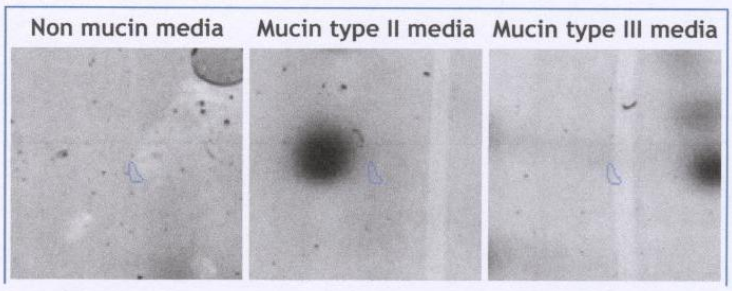


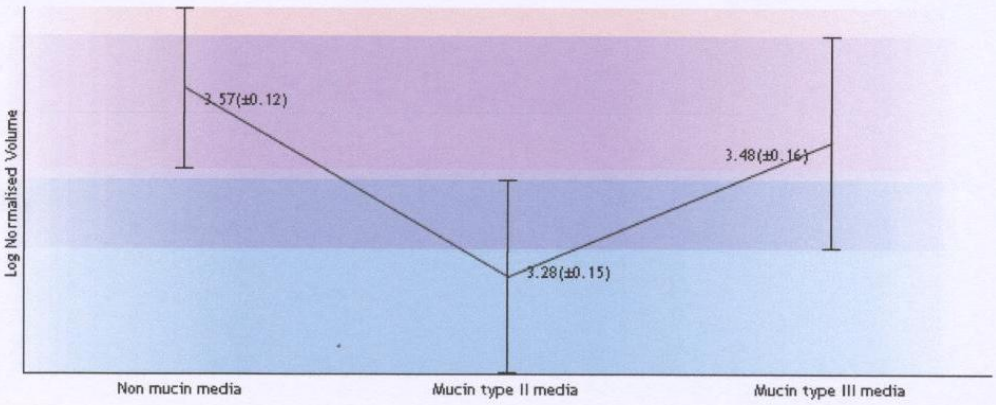
Identifier 2352

Position (496, 1470)

Notes

- p value spots
- Power spots
- spots of interest



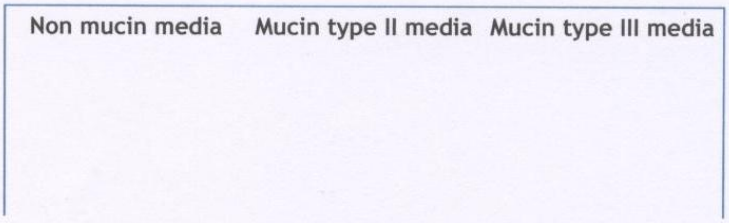


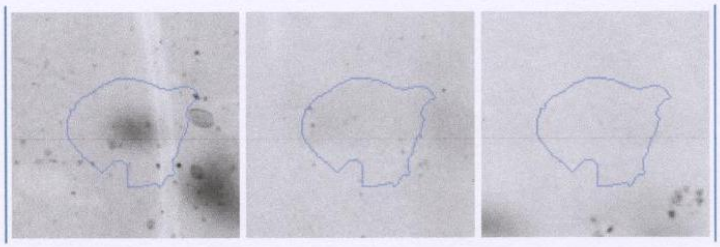
Identifier 2410

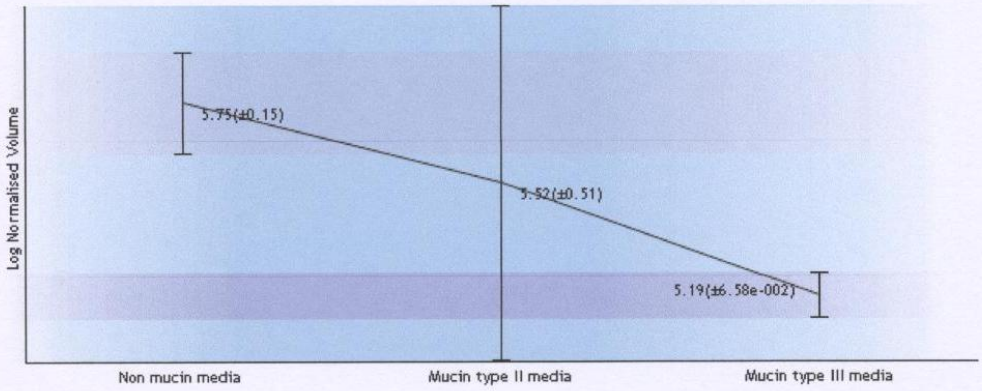
Position (795, 1507)

Notes

- fold change of 3.0 or more
- p value spots
- Power spots
- spots of interest







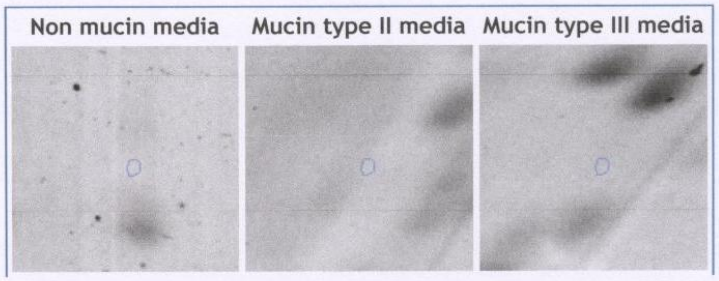
Identifier 1579

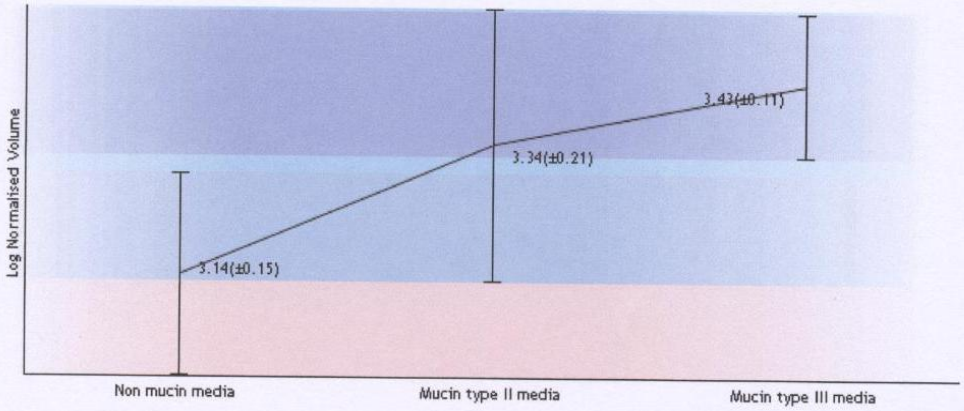
Position (1121, 960)

Notes

- p value spots
- Power spots
- spots of interest

Non mucin media	Mucin type II media	Mucin type III media



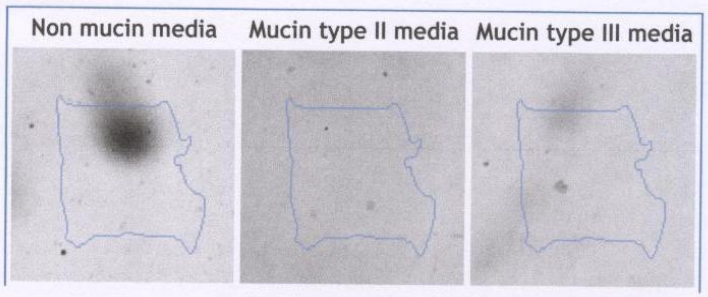


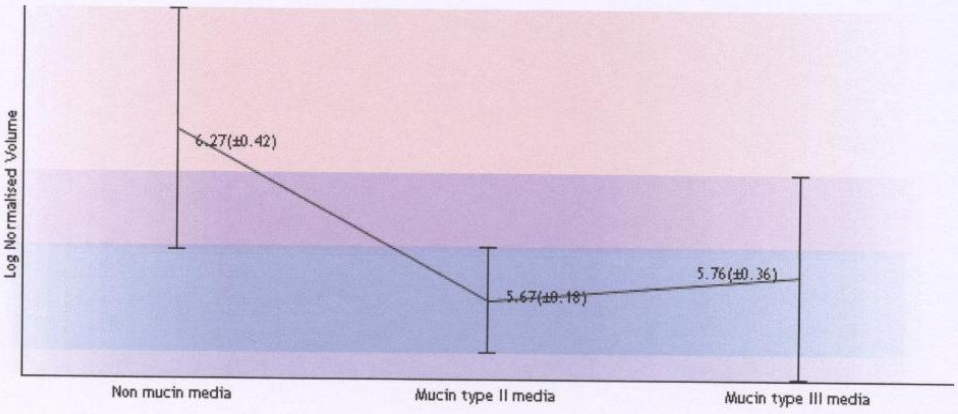
Identifier 1342

Position (1533, 827)

Notes

- fold change of 3.0 or more
- p value spots
- Power spots
- spots of interest



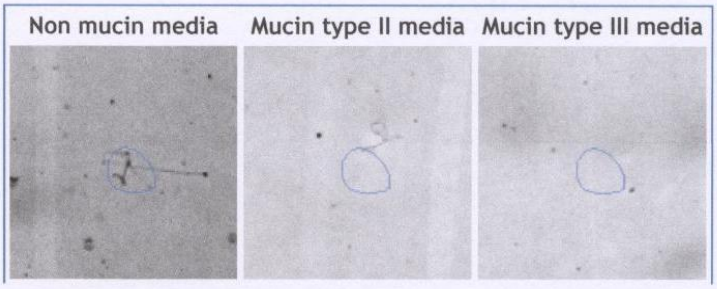


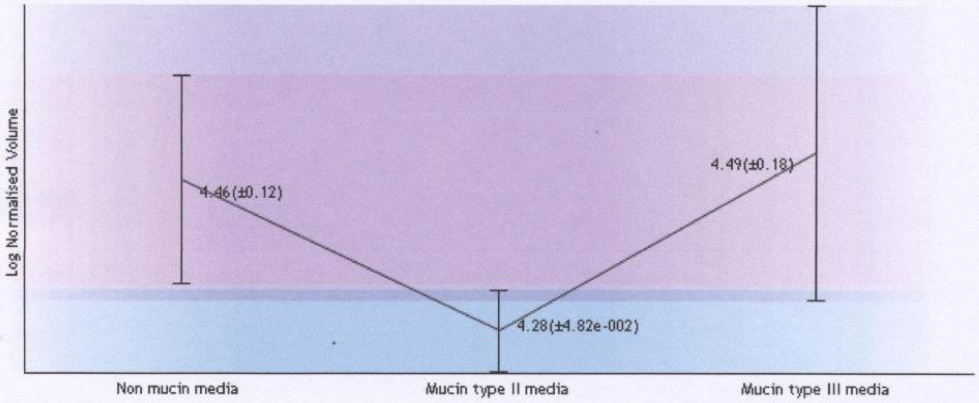
Identifier 3060

Position (1918, 1933)

Notes

- p value spots
- Power spots
- spots of interest



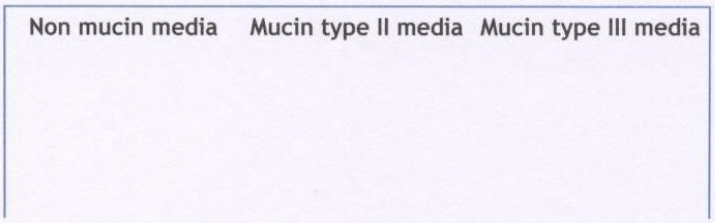


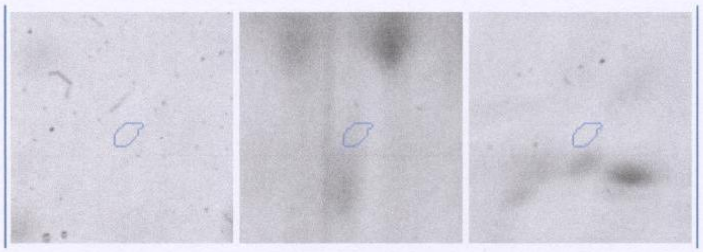
Identifier 813

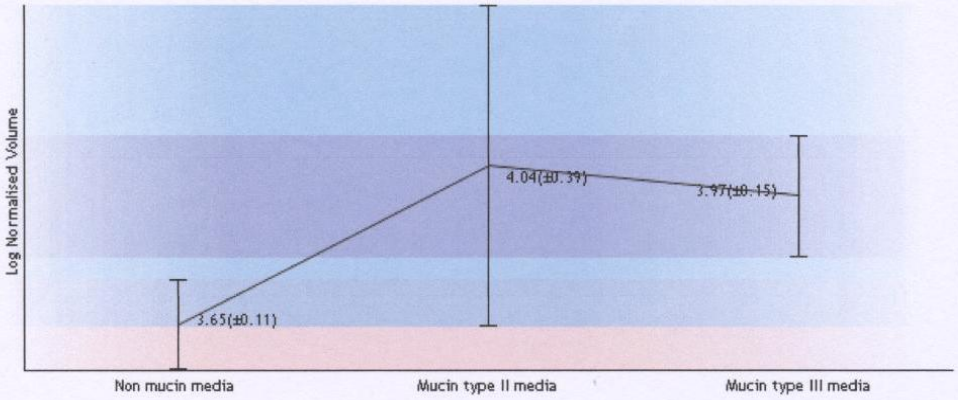
Position (1270, 552)

Notes

- fold change of 3.0 or more
- p value spots
- Power spots
- spots of interest







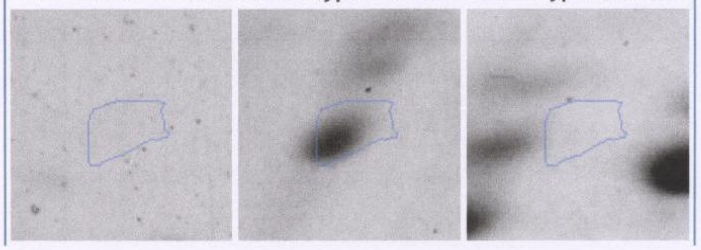
Identifier 1922

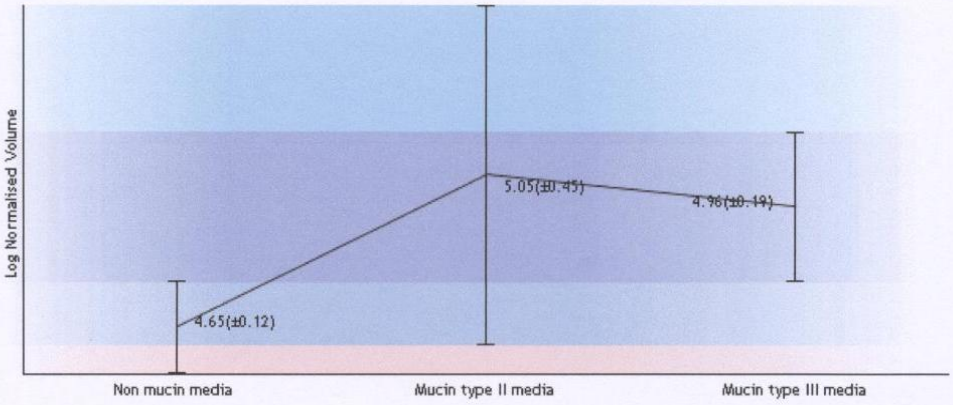
Position (1171, 1158)

Notes

- fold change of 3.0 or more
- p value spots
- spots of interest

Non mucin media	Mucin type II media	Mucin type III media



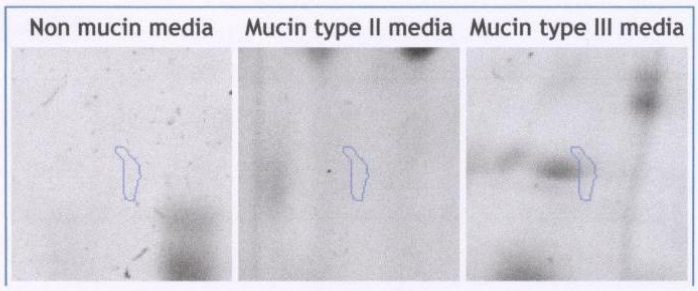


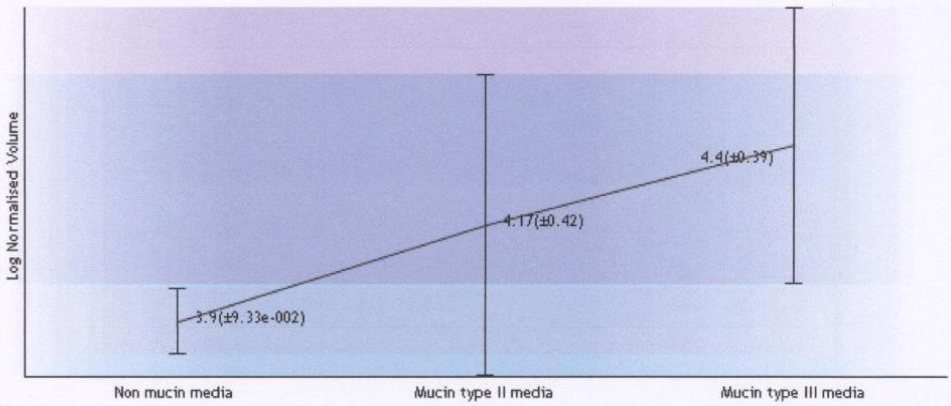
Identifier 872

Position (1314, 579)

Notes

- fold change of 3.0 or more
- p value spots
- spots of interest



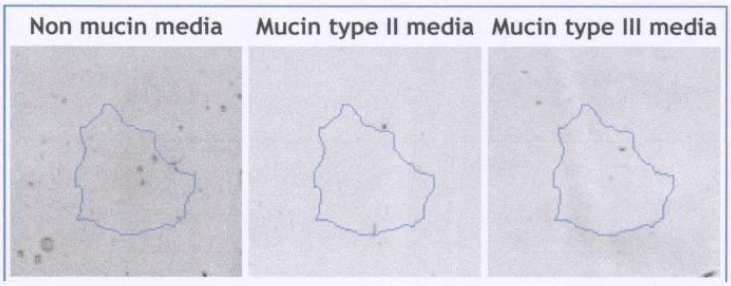


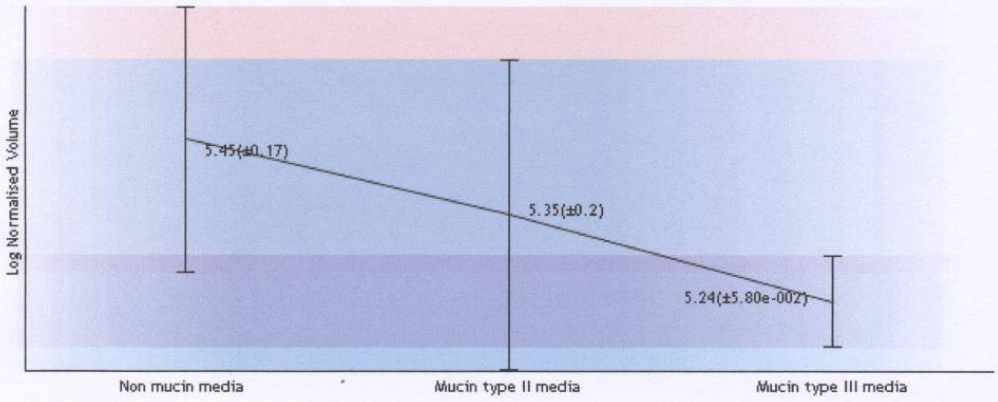
Identifier 3244

Position (1067, 2075)

Notes

- p value spots
- spots of interest



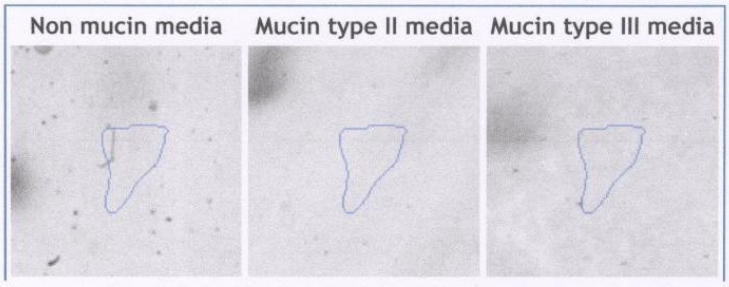


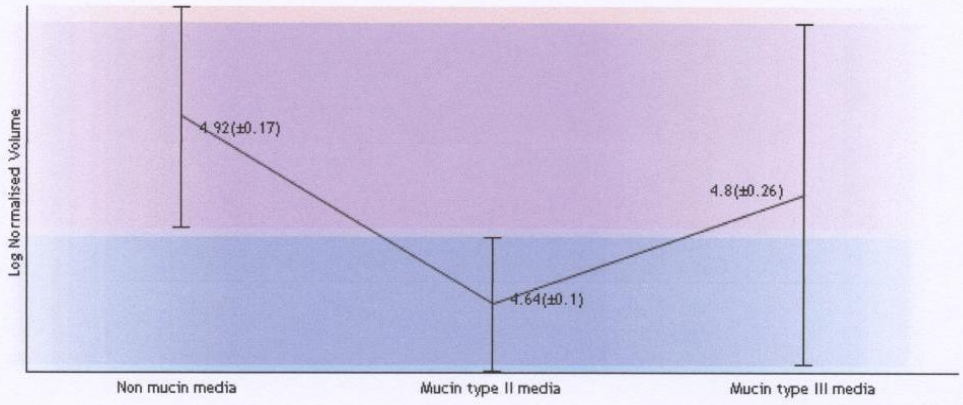
Identifier 2083

Position (1523, 1251)

Notes

- p value spots
- spots of interest



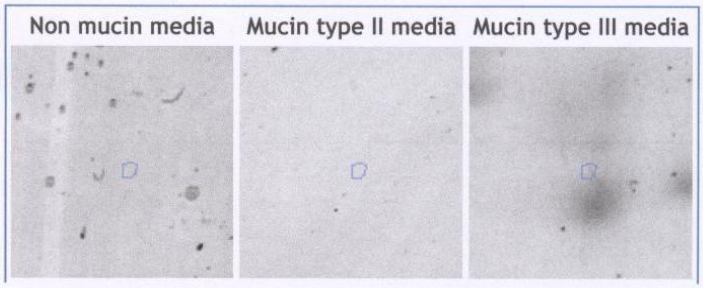


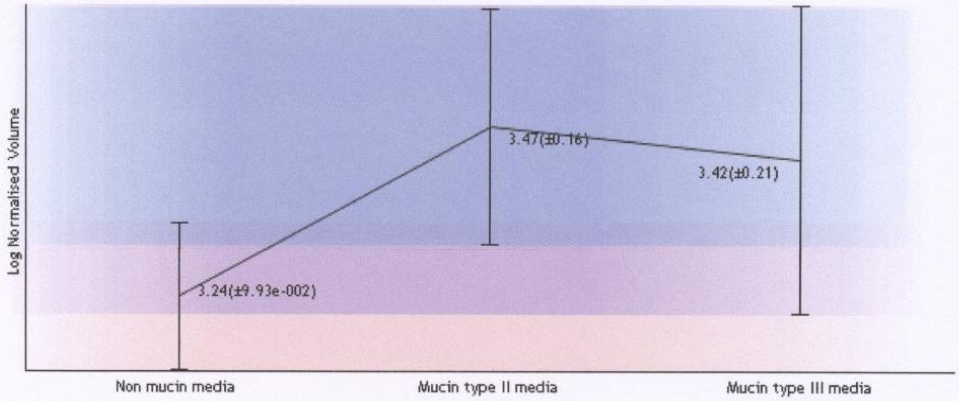
Identifier 1902

Position (1923, 1144)

Notes

- p value spots
- spots of interest



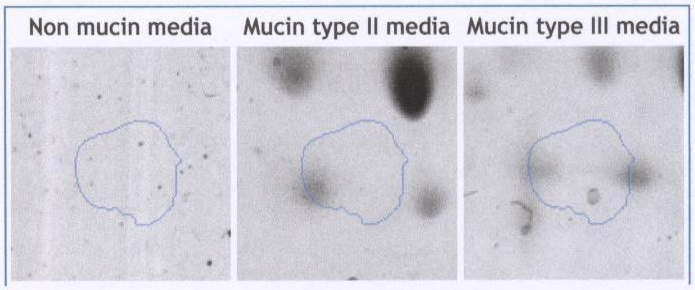


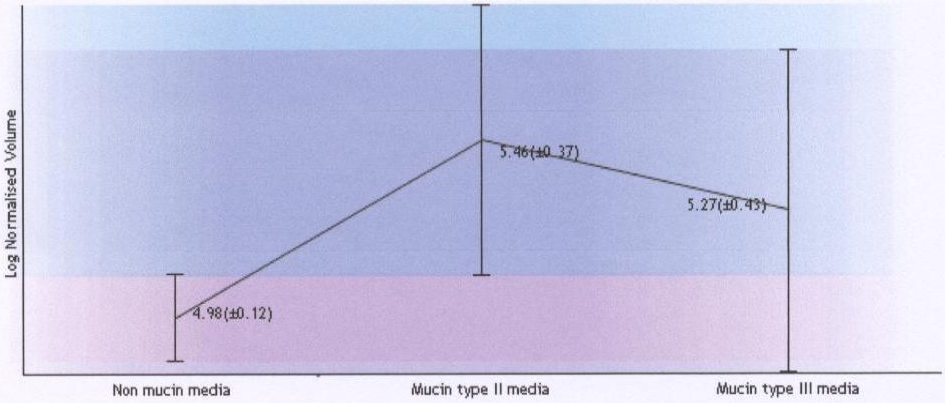
Identifier 285

Position (1797, 259)

Notes

- fold change of 3.0 or more
- p value spots
- spots of interest



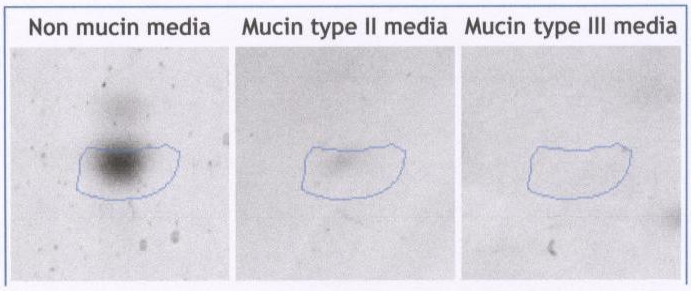


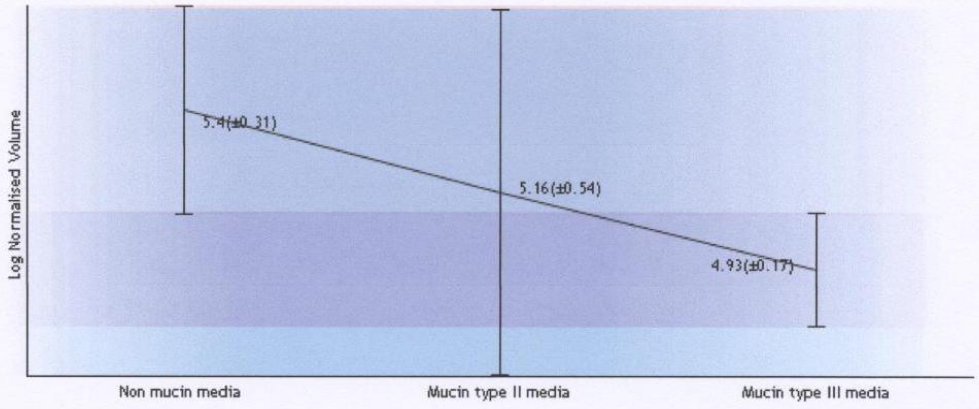
Identifier 1807

Position (1461, 1079)

Notes

- fold change of 3.0 or more
- p value spots
- spots of interest



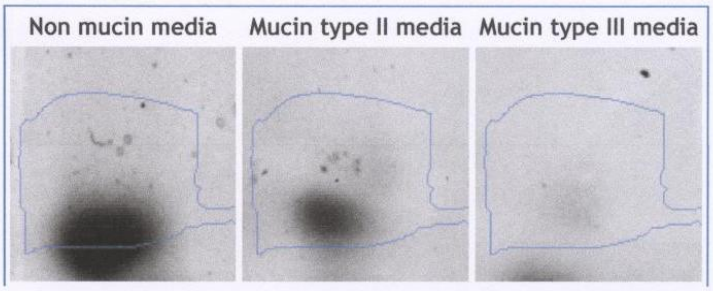


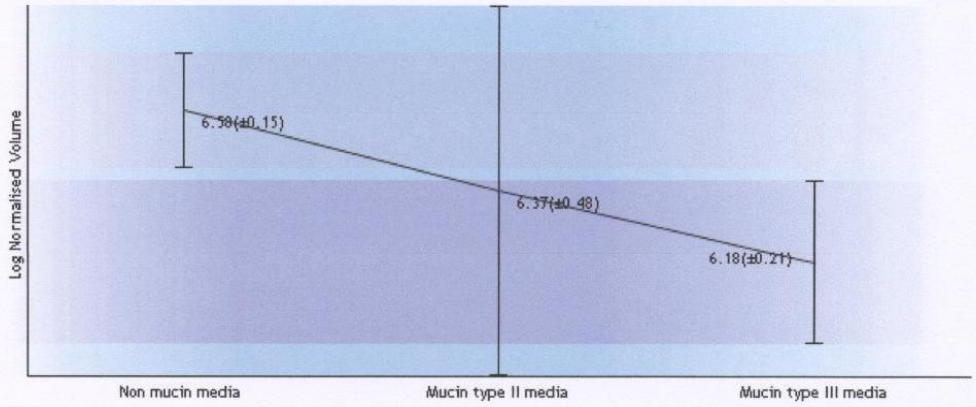
Identifier 2712

Position (923, 1717)

Notes

- p value spots
- spots of interest



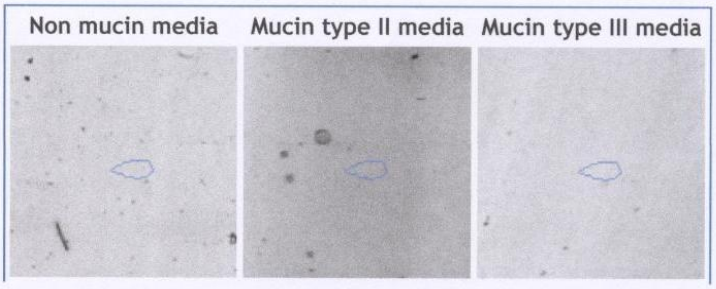


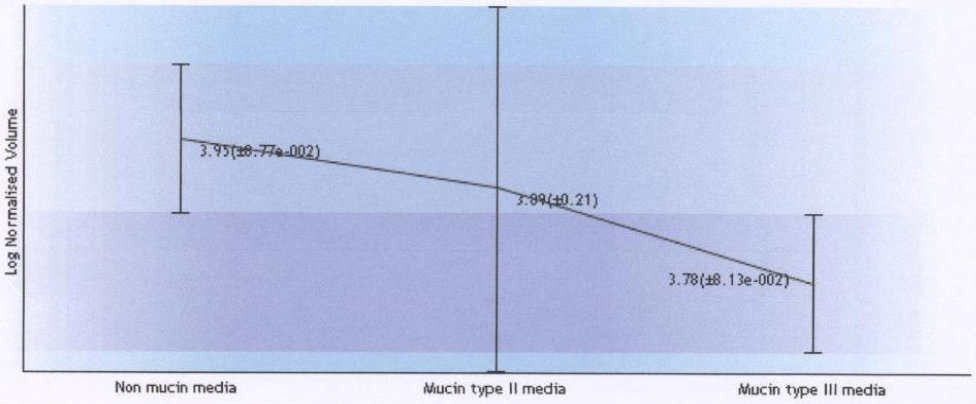
Identifier 835

Position (666, 565)

Notes

- p value spots
- spots of interest





Appendix G

G1 Details on MS analysis

Bruker Daltonics Mass spectrometer that worked based on the principle of electrospray ionisation was used for fragmenting, ionising and identifying peptides from the protein spots of interest. The liquid sample which passed through an emitter was ionised to form a liquid jet under high voltage electricity. The liquid then radially dispersed to form small and highly charged liquid droplets due to Coulomb repulsion. Following nebulisation, the ions were dried and desorbed under the influence of the same high potential electrostatic field.

The generation of ions

The three important steps in the generation of ions were nebulisation, desolvation and ion evaporation. The sample containing the trypsinised proteins was first separated using the LC column in the Dionex Ultimate 3000 system and the droplets were fed into the mass spectrometer through the needle at a flow rate of 3 µL/min. The droplets enter the spray chamber from the tip of the needle where the force of Coulomb repulsion exceeds the surface tension causing the droplets to explode into smaller droplets that undergo evaporation. During the process of evaporation, the droplets become highly charged and get ionised. When the surface tension on the surface of the droplets is exceeded by the ionic field, ions are emitted into the gas phase. Peptides containing amino acids either accept or donate protons based on their charge. A similar effect can be observed in the buffers and solvents that also ionise in the presence of an electric field. pH of the buffers and solvents can have a great effect on the peptide signals detected. The ionisation process can also be affected by the conductivity of solutions and their flow rate.

Parts of the Ion trap mass spectrometer

The spray shield forms a protective covering that prevents the deposition of contaminants, buffer salts and other impurities. The transfer capillary plays an important role in the declustering and desolvation of ions. It acts as the first barrier in the vaccum system. The skimmer forms the main part of the next vaccum stage where it helps to remove dried gas and other solvent molecules. The ions that pass

through the transfer capillary and the skimmer enter the dual octopoles that focus, slow down or accelerate them for efficient ion trapping.

Ion trap system

The first step in the trapping of ions involves the generation of a potential well which acts as the site for accumulation of ions. The end cap generates a repelling potential that enhances the trapping efficiency. The ions collide with helium thereby slowing them down and oscillating around the centre of the trap. The ions focussed in the centre of the trap are ejected by applying an increased RF amplitude and auxillary RF which cause each m/z to come into resonance. The ions accumulated in the 3D multipolar field were ejected by increasing the RF amplitudes and the response was measured at the detector.

The conversion dynode emitted electrons when hit by ions. These electrons were accelerated onto an electron multiplier that converted the electrons into a resulting current that was measured as the MS spectrum.

Gases used in Ion traps

One rough pump and two turbomolecular pumps were used to create four stages of vaccum within the system (Refer to drawing). Helium gas was used for cooling the ions during trapping and also as a collision gas for fragmentation. Nitrogen gas was used generally for keeping the system clean, nebulising samples and drying liquid droplets.

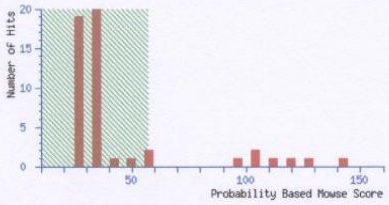
G2 Mass spectrometric analysis data for differentially expressed spots *E. cancerogenus*

Mascot Search Results

```
User          : Lakshmy Manickan
Email         : lakshmy.manickan@unn.ac.uk
Search title  : SSP 2302
MS data file  : D:\Data\Lakshmy\120808\140808\SSP2302_RE13_01_281.d\SSP2302_RE13_01_281.mgf
Database      : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy      : Bacteria (Eubacteria) (5690016 sequences)
Timestamp     : 5 Jan 2010 at 11:11:44 GMT
Protein hits  : gi|126496     RecName: Full=Lysostaphin; AltName: Full=Glycyl-glycine endopeptidase; Flags: Precursor
                gi|261342624 transcription elongation factor GreA [Enterobacter cancerogenus ATCC 35316]
                gi|16762061  transcription elongation factor GreA [Salmonella enterica subsp. enterica serovar Typhi str. CT18]
                gi|37528357  transcription elongation factor GreA [Photorhabdus luminescens subsp. laumondii TTO1]
                gi|226329694 hypothetical protein PROPEN_03606 [Proteus penneri ATCC 35198]
                gi|238750117 Transcription elongation factor greA [Yersinia rohdei ATCC 43380]
                gi|269137754 transcription elongation factor [Edwardsiella tarda EIB202]
                gi|238757567 Transcription elongation factor greA [Yersinia aldovae ATCC 35236]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|126496 **Mass:** 51656 **Score:** 143 **Queries matched:** 2 **emPAI:** 0.13
RecName: Full=Lysostaphin; AltName: Full=Glycyl-glycine endopeptidase; Flags: Precursor
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 102	819.8730	1637.7314	1637.7897	-0.0582	0	42	1.7	1	K.SESASFTPNTDIITR.T
☑ 182	1007.4620	2012.9094	2012.9336	-0.0241	0	101	2e-06	1	R.MVNSFSNSTAQDPMPFLK.S

Proteins matching the same set of peptides:
gi|153047 **Mass:** 42213 **Score:** 143 **Queries matched:** 2
lysostaphin (ttg start codon) [Staphylococcus simulans]
gi|3287967 **Mass:** 53058 **Score:** 143 **Queries matched:** 2
RecName: Full=Lysostaphin; AltName: Full=Glycyl-glycine endopeptidase; Flags: Precursor

---

2. gi|261342624 **Mass:** 17799 **Score:** 125 **Queries matched:** 3 **emPAI:** 0.42
transcription elongation factor GreA [Enterobacter cancerogenus ATCC 35316]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 72	656.3350	1310.6554	1310.7306	-0.0752	0	61	0.022	1	K.QNLISVNSPIAR.G
☑ 74	667.8120	1333.6094	1333.6765	-0.0671	0	40	2.9	1	R.TPGGEVEYEIIK.V
☑ 145	919.9400	1837.8654	1837.9349	-0.0695	1	24	1.1e+02	1	R.TPGGEVEYEIIKVEYL.-

---

3. gi|16762061 **Mass:** 17703 **Score:** 120 **Queries matched:** 3 **emPAI:** 0.69
transcription elongation factor GreA [Salmonella enterica subsp. enterica serovar Typhi str. CT18]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 70	655.3790	1308.7434	1308.7513	-0.0079	0	(57)	0.059	1	R.RPEIIAAIAEAR.E
☑ 71	437.2600	1308.7582	1308.7513	0.0068	0	58	0.039	1	R.RPEIIAAIAEAR.E
72	656.3350	1310.6554	1310.7306	-0.0752	0	61	0.022	1	K.QNLISVNSPIAR.G

Proteins matching the same set of peptides:
gi|24114470 **Mass:** 17688 **Score:** 120 **Queries matched:** 3
transcription elongation factor GreA [Shigella flexneri 2a str. 301]

Mascot Search Results

Protein View

Match to: **gi|261342624** Score: **125**
transcription elongation factor GreA [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\140808\SSP2302_RE13_01_281.d\SSP2302_RE13_01_281.mgf

Nominal mass (M_r): **17799**; Calculated pI value: **4.82**
NCBI BLAST search of gi|261342624 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **17%**

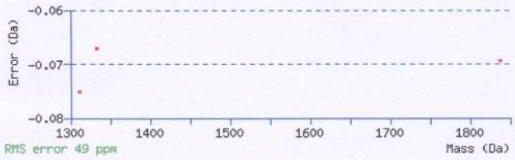
Matched peptides shown in **Bold Red**

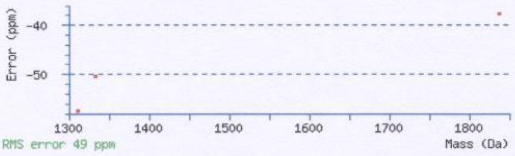
```
  1 MQAIPMTLRG AEKLREELDF LKSVRRPEII AAIADAREHG DLKENAEYHA
 51 AREQQGFCEG RIKDIEAKLS NAQVIDITKM PNNGRVIFGA TVTVLNLDND
101 EEQKYRIVGD DEADFKQNLI SVNSPIARGL IGKEQDDVVV IRTPGGEVEY
151 EIIKVEYL
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence	
117 - 128	656.3350	1310.6554	1310.7306	-0.0752	0	K.QNLISVNSPIAR.G	(Ions score 61)
143 - 154	667.8120	1333.6094	1333.6765	-0.0671	0	R.TPGGEVEYEIIK.V	(Ions score 40)
143 - 158	919.9400	1837.8654	1837.9349	-0.0695	1	R.TPGGEVEYEIIKVEYL.-	(Ions score 24)






```
LOCUS       ZP_05970482              158 aa            linear   BCT 15-OCT-2009
DEFINITION  transcription elongation factor GreA [Enterobacter cancerogenus
            ATCC 35316].
ACCESSION   ZP_05970482
VERSION     ZP_05970482.1  GI:261342624
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000035.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 158)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 158)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000035.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
```

```
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..158
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..158
                     /product="transcription elongation factor GreA"
                     /calculated_mol_wt=17621
     Region          1..157
                     /region_name="greA"
                     /note="transcription elongation factor GreA; Reviewed;
                     PRK00226"
                     /db_xref="CDD:134177"
     Region          1..74
                     /region_name="GreA_GreB_N"
                     /note="N-terminal domain; pfam03449"
                     /db_xref="CDD:112274"
     Region          80..157
                     /region_name="GreA_GreB"
                     /note="C-terminal domain; pfam01272"
                     /db_xref="CDD:110285"
     CDS             1..158
                     /gene="greA"
                     /locus_tag="EcanA3_020100019305"
                     /coded_by="complement(NZ_ABWM02000035.1:306544..307020)"
                     /note="COG0782 Transcription elongation factor"
                     /transl_table=11
                     /db_xref="CDD:134177"
```

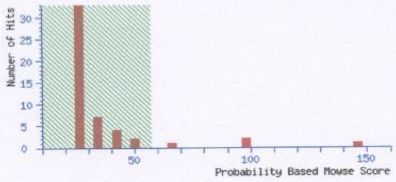
Mascot: http://www.matrixscience.com/

Mascot Search Results

User : Lakshmy Manickan
Email : lakshmy.manickan@unn.ac.uk
Search title : SSP 3002
MS data file : D:\Data\Lakshmy\120808\260808\SSP 3002_RG16_01_396.d\SSP 3002_RG16_01_396.mgf
Database : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy : Bacteria (Eubacteria) (5690016 sequences)
Timestamp : 5 Jan 2010 at 11:13:53 GMT
Protein hits : gi|42543180 Chain A, Ige Fv Spe7 Complexed With A Recombinant Thioredoxin
gi|15802068 hypothetical protein Z2676 [Escherichia coli O157:H7 EDL933]
gi|238920001 hypothetical protein NT01EI_2105 [Edwardsiella ictaluri 93-146]
gi|54310594 putative thioredoxin [Photobacterium profundum SS9]

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As Peptide Summary Help
Significance threshold p< 0.05 Max. number of hits AUTO
Standard scoring MudPIT scoring Ions score or expect cut-off 0 Show sub-sets 0
Show pop-ups Suppress pop-ups Sort unassigned Decreasing Score Require bold red
Select All Select None Search Selected Error tolerant

1. gi|42543180 Mass: 13560 Score: 146 Queries matched: 4 emPAI: 0.96
Chain A, Ige Fv Spe7 Complexed With A Recombinant Thioredoxin
Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
128	607.9250	1820.7532	1820.8866	-0.1334	0	(21)	1.8e+02	1	K.MIAPILDEIADEYQGK.L + Oxidation (M)
129	607.9690	1820.8852	1820.8866	-0.0014	0	(61)	0.022	1	K.MIAPILDEIADEYQGK.L + Oxidation (M)
130	911.4580	1820.9014	1820.8866	0.0149	0	113	1.3e-07	1	K.MIAPILDEIADEYQGK.L + Oxidation (M)
131	913.7260	2738.1562	2738.3183	-0.1621	1	35	11	1	R.YDLVGPCKMIAPILDEIADEYQGK.L

2. gi|15802068 Mass: 13044 Score: 101 Queries matched: 2 emPAI: 0.26
hypothetical protein Z2676 [Escherichia coli O157:H7 EDL933]
Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
81	724.9020	1447.7894	1447.7744	0.0150	0	32	18	1	R.QIAENPILLYMK.G + Oxidation (M)
91	782.4400	1562.8654	1562.8093	0.0562	0	69	0.0033	1	R.FAYVDILQNPDIR.A

Proteins matching the same set of peptides:
gi|16760480 Mass: 13074 Score: 101 Queries matched: 2
hypothetical protein STY1689 [Salmonella enterica subsp. enterica serovar Typhi str. CT18]
gi|50120857 Mass: 13245 Score: 101 Queries matched: 2
hypothetical protein ECA1927 [Pectobacterium atrosepticum SCRI1043]
gi|74312019 Mass: 13044 Score: 101 Queries matched: 2
hypothetical protein SSON_1502 [Shigella sonnei Ss046]
gi|123442416 Mass: 13060 Score: 101 Queries matched: 2
hypothetical protein YE2155 [Yersinia enterocolitica subsp. enterocolitica 8081]
gi|152970544 Mass: 13068 Score: 101 Queries matched: 2
hypothetical protein KPN_01992 [Klebsiella pneumoniae subsp. pneumoniae MGH 78578]
gi|156934186 Mass: 12933 Score: 101 Queries matched: 2
hypothetical protein ESA_02013 [Enterobacter sakazakii ATCC BAA-894]
gi|157145914 Mass: 13074 Score: 101 Queries matched: 2
hypothetical protein CKO_01665 [Citrobacter koseri ATCC BAA-895]
gi|161503474 Mass: 13002 Score: 101 Queries matched: 2
hypothetical protein SARI_01548 [Salmonella enterica subsp. arizonae serovar 62:z4,z23:--]
gi|170768703 Mass: 13030 Score: 101 Queries matched: 2
putative monothiol glutaredoxin [Escherichia albertii TW07627]
gi|188533942 Mass: 12734 Score: 101 Queries matched: 2
hypothetical protein ETA_18040 [Erwinia tasmaniensis Et1/99]
gi|227111796 Mass: 13212 Score: 101 Queries matched: 2
hypothetical protein PcarbP_02467 [Pectobacterium carotovorum subsp. brasiliensis PBR1692]
gi|237731383 Mass: 13089 Score: 101 Queries matched: 2

conserved hypothetical protein [Citrobacter sp. 30_2]
gi|238749425 **Mass:** 13318 **Score:** 101 **Queries matched:** 2
Glutaredoxin-4 [Yersinia rohdei ATCC 43380]
gi|238782581 **Mass:** 13289 **Score:** 101 **Queries matched:** 2
Glutaredoxin-4 [Yersinia bercovieri ATCC 43970]
gi|238787142 **Mass:** 13288 **Score:** 101 **Queries matched:** 2
Glutaredoxin-4 [Yersinia frederiksenii ATCC 33641]
gi|238792292 **Mass:** 13262 **Score:** 101 **Queries matched:** 2
Glutaredoxin-4 [Yersinia intermedia ATCC 29909]
gi|241913123 **Mass:** 15206 **Score:** 101 **Queries matched:** 2
Chain A, Structure Of E. Coli Monothiol Glutaredoxin Grx4 Homodimer
gi|242239124 **Mass:** 12992 **Score:** 101 **Queries matched:** 2
glutaredoxin-like protein [Dickeya dadantii Ech703]
gi|251789329 **Mass:** 13316 **Score:** 101 **Queries matched:** 2
glutaredoxin-like protein [Dickeya zeae Ech1591]
gi|253688763 **Mass:** 13228 **Score:** 101 **Queries matched:** 2
glutaredoxin-like protein [Pectobacterium carotovorum subsp. carotovorum PC1]
gi|258635113 **Mass:** 12512 **Score:** 101 **Queries matched:** 2
glutaredoxin-like protein [Pantoea sp. At-9b]
gi|259908543 **Mass:** 12733 **Score:** 101 **Queries matched:** 2
Putative glutaredoxin protein YdhD [Erwinia pyrifoliae Ep1/96]
gi|260597772 **Mass:** 12927 **Score:** 101 **Queries matched:** 2
Glutaredoxin-4 [Cronobacter turicensis]
gi|261339586 **Mass:** 13045 **Score:** 101 **Queries matched:** 2
hypothetical protein EcanA3_03947 [Enterobacter cancerogenus ATCC 35316]
gi|261821927 **Mass:** 13259 **Score:** 101 **Queries matched:** 2
glutaredoxin-like protein [Pectobacterium wasabiae WPP163]
gi|269139028 **Mass:** 12773 **Score:** 101 **Queries matched:** 2
hypothetical protein ETAE_1679 [Edwardsiella tarda EIB202]
gi|271500215 **Mass:** 13334 **Score:** 101 **Queries matched:** 2
glutaredoxin-like protein [Dickeya dadantii Ech586]

---

3. gi|238920001 **Mass:** 12760 **Score:** 101 **Queries matched:** 2 **emPAI:** 0.27
hypothetical protein NT01EI_2105 [Edwardsiella ictaluri 93-146]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
81	724.9020	1447.7894	1447.7744	0.0150	0	32	18	1	R.QIAENPIILYMK.G + Oxidation (M)
91	782.4400	1562.8654	1562.8093	0.0562	0	69	0.0033	1	R.FAYVDILQNPDIR.A

---

4. gi|54310594 **Score:** 65 **Queries matched:** 3
putative thioredoxin [Photobacterium profundum SS9]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
128	607.9250	1820.7532	1820.8866	-0.1334	0	(19)	3.2e+02	2	K.MIAPILDEIANEYEGK.V + Oxidation (M)
129	607.9690	1820.8852	1820.8866	-0.0014	0	(58)	0.044	2	K.MIAPILDEIANEYEGK.V + Oxidation (M)
130	911.4580	1820.9014	1820.8866	0.0149	0	65	0.0079	2	K.MIAPILDEIANEYEGK.V + Oxidation (M)

Proteins matching the same set of peptides:
gi|84393642 **Score:** 65 **Queries matched:** 3
gi|90414228 **Score:** 65 **Queries matched:** 3

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	70	679.3270	1356.6394	1356.7976	-0.1582	2	45	1	1	LNDIDSILKAKK
☑	55	515.2470	1028.4794	1028.5866	-0.1071	0	33	18	1	VLDELLTAR
☑	65	632.3680	1262.7214	1263.6320	-0.9105	1	31	23	1	REQVQNATYR
☑	71	681.1670	2040.4792	2040.1367	0.3424	1	29	43	1	NIVSSTELHQNVIKAFIK
☑	52	492.3030	982.5914	983.5916	-1.0002	1	29	32	1	KIATWLPR
☑	58	582.1380	1743.3922	1743.8540	-0.4618	0	29	50	1	NLPPLYNSQPNSASSR
☑	198	737.6880	2210.0422	2211.0598	-1.0177	0	28	33	1	EHPRPPVMMSTASQMLVER + Oxidation (M)
☑	136	615.8910	1844.6512	1845.0472	-0.3960	1	26	44	1	ILNDSIAYIRSLAQLR
☑	225	1164.4980	3490.4722	3489.7207	0.7515	2	26	77	1	DYGIFLVHDASPVKMGGMPVLPMTGSRQESVR + Oxidation (M)
☑	77	701.4550	2101.3432	2101.0361	0.3070	1	26	1.1e+02	1	CAGLADGPRSSVVLIDEVDK
☑	202	742.3760	2224.1062	2223.1681	0.9380	0	26	62	1	MLTGTATLPSTAAAQPPAQIQR
☑	44	465.7930	1394.3572	1393.6403	0.7168	0	26	1.5e+02	1	AAVIEAMAQCMAK
☑	121	900.0780	2697.2122	2698.3021	-1.0899	0	25	1.2e+02	1	QIGSGQEAAPAFSPPSTLSMQAPSPAR + Oxidation (M)
☑	193	1084.4090	3250.2052	3250.3690	-0.1639	1	25	92	1	MMFRGFGLAMPSASYNDDWMDGVYQMAK + 2 Oxidation (M)
☑	122	904.2980	1806.5814	1805.8180	0.7634	2	25	1e+02	1	VVGDFLDDNDNREKGN
☑	132	609.5150	1825.5232	1824.8135	0.7097	0	24	66	1	SMGGNDTPAVGFAMGIER + Oxidation (M)
☑	106	847.9540	2540.8402	2540.3737	0.4664	1	24	1.6e+02	1	IELIKDYLIHPEGSVLISTGDTK
☑	241	905.5410	2713.6012	2714.4391	-0.8380	1	24	69	1	RIGLSAHDFAQLLGADEVILTYANK
☑	109	577.9310	1730.7712	1729.8999	0.8713	0	24	1.1e+02	1	NDSAIVQALEALFPSR
☑	144	951.8300	2852.4682	2851.4279	1.0402	2	24	1.6e+02	1	KFGIVPEFEDEREEIQEIEQFIK
☑	73	692.3460	1382.6774	1383.6704	-0.9929	0	23	1.3e+02	1	MSQSSALASFLDK
☑	218	767.1490	2298.4252	2299.2001	-0.7749	1	23	84	1	FRFADTLGILDPFATFEAIR
☑	152	974.8230	2921.4472	2920.3735	1.0737	2	23	1.8e+02	1	KEPMIADLARYENQSITGFFCASQK + Oxidation (M)
☑	45	931.6700	930.6627	931.4611	-0.7983	0	23	1.7e+02	1	AETSQTPAK
☑	89	771.6520	2311.9342	2311.1466	0.7876	2	23	1.9e+02	1	MHAQASLFCFRAARIAFASR
☑	124	906.2870	2715.8392	2716.2477	-0.4086	2	23	1.7e+02	1	DPYSQGARAGDKFDPYTQGTNATTR
☑	53	499.3350	996.6554	996.4624	0.1930	0	23	1.4e+02	1	SQAALDNAAH
☑	161	991.6440	2971.9102	2972.5351	-0.6249	2	23	1.8e+02	1	LKFIDVVIVDTAGRLHVDEAMMDEIK + Oxidation (M)
☑	221	1158.1100	3471.3082	3470.7359	0.5723	1	22	1.9e+02	1	HNGVDAVAMTRAVGQALVNGGVVSGASTLTMQTAR + 2 Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|261339586** Score: **101**
hypothetical protein EcanA3_03947 [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\260808\SSP 3002_RG16_01_396.d\SSP 3002_RG16_01_396.mgf

Nominal mass (M_r): **13045**; Calculated pI value: **4.87**
NCBI BLAST search of gi|261339586 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **21%**

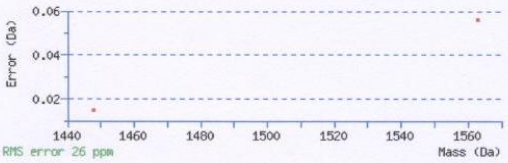
Matched peptides shown in **Bold Red**

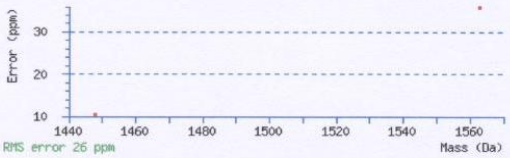
```
  1 MSTTIEKIQR QIAENPILLY MKGSPKLPSC GFSAQAVQAL SACGERFAYV
 51 DILQNPDIRA ELPKYANWPT FPQLWIDGEL VGGCDILIEM YQRGELQQLI
101 KETAAKYKTE EPGAE
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
11 - 22	724.9020	1447.7894	1447.7744	0.0150	0	R.QIAENPILLYMK.G Oxidation (M) (Ions score 32)
47 - 59	782.4400	1562.8654	1562.8093	0.0562	0	R.FAYVDILQNPDIR.A (Ions score 69)






```
LOCUS       ZP_05967444              115 aa            linear   BCT 15-OCT-2009
DEFINITION  hypothetical protein EcanA3_03947 [Enterobacter cancerogenus ATCC
            35316].
ACCESSION   ZP_05967444
VERSION     ZP_05967444.1  GI:261339586
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000005.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 115)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 115)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000005.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
```

```
          cancerogenus, not the sequenced strain), is a member of the
          Proteobacteria division of the domain bacteria and has been
          isolated from human feces. The sequenced strain was obtained from
          ATCC (ATCC 35316).

          We have performed one round of automated sequence improvement
          (pre-finishing), along with manual improvement that includes
          breaking apart any mis-assembly, and making manual joins where
          possible. Manual edits also are made where the consensus appears to
          be incorrect. All low quality data on the ends of contigs is
          removed. Contigs are ordered and oriented where possible.

          This work was supported by the National Human Genome Research
          Institute and the NIH Roadmap Human Microbiome Project
          [http://nihroadmap.nih.gov/hmp/].

          This is a reference genome for the Human Microbiome Project.
          This project is co-owned with the Human Microbiome Project DACC.
          Genome Coverage: 29.7x
          Sequencing Technology: 454.
          Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..115
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..115
                     /product="hypothetical protein"
                     /calculated_mol_wt=12748
     Region          8..97
                     /region_name="GRX_PICOT_like"
                     /note="Glutaredoxin (GRX) family, PKC-interacting cousin
                     of TRX (PICOT)-like subfamily; composed of PICOT and
                     GRX-PICOT-like proteins. The non-PICOT members of this
                     family contain only the GRX-like domain, whereas PICOT
                     contains an N-terminal TRX-like domain...; cd03028"
                     /db_xref="CDD:48577"
     Site            order(22,30..32,70..71)
                     /site_type="other"
                     /note="putative GSH binding site"
                     /db_xref="CDD:48577"
     Site            order(30,33)
                     /site_type="other"
                     /note="catalytic residues"
                     /db_xref="CDD:48577"
     CDS             1..115
                     /locus_tag="EcanA3_020100003947"
                     /coded_by="NZ_ABWM02000005.1:254505..254852"
                     /note="COG0278 Glutaredoxin-related protein"
                     /transl_table=11
                     /db_xref="CDD:138199"
```

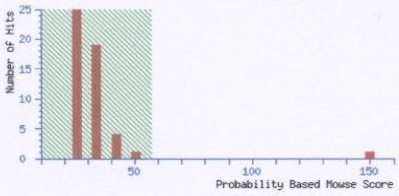
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User          : Lakshmy Manickan
Email         : lakshmy.manickan@unn.ac.uk
Search title  : SSP 4008
MS data file  : D:\Data\meng\20080728\SSP4008_RB7_01_222.d\SSP4008_RB7_01_222.mgf
Database      : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy      : Bacteria (Eubacteria) (5690016 sequences)
Timestamp     : 5 Jan 2010 at 11:17:21 GMT
Protein hits  : gi|15803106 autonomous glycyl radical cofactor GrcA [Escherichia coli O157:H7 EDL933]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|15803106 **Mass:** 14283 **Score:** 150 **Queries matched:** 4 **emPAI:** 0.53
autonomous glycyl radical cofactor GrcA [Escherichia coli O157:H7 EDL933]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 47	445.7760	889.5374	889.5021	0.0353	0	38	5.7	1	K.YPQLTIR.V
☑ 110	478.9540	1433.8402	1433.7739	0.0663	0	(37)	6.2	1	R.VEGGQHLNVNVLR.R
☑ 111	717.9630	1433.9114	1433.7739	0.1375	0	39	3.3	1	R.VEGGQHLNVNVLR.R
☑ 184	976.0110	1950.0074	1949.9370	0.0704	0	75	0.00082	1	K.AANDDLLNSFWLLDSEK.G

Proteins matching the same set of peptides:
gi|15832699 **Mass:** 14333 **Score:** 150 **Queries matched:** 4
autonomous glycyl radical cofactor GrcA [Escherichia coli O157:H7 str. Sakai]
gi|16761504 **Mass:** 14393 **Score:** 150 **Queries matched:** 4
autonomous glycyl radical cofactor GrcA [Salmonella enterica subsp. enterica serovar Typhi str. CT18]
gi|82778007 **Mass:** 14375 **Score:** 150 **Queries matched:** 4
autonomous glycyl radical cofactor GrcA [Shigella dysenteriae Sd197]
gi|110642741 **Mass:** 14317 **Score:** 150 **Queries matched:** 4
autonomous glycyl radical cofactor GrcA [Escherichia coli 536]
gi|146312705 **Mass:** 14385 **Score:** 150 **Queries matched:** 4
autonomous glycyl radical cofactor GrcA [Enterobacter sp. 638]
gi|152971433 **Mass:** 14350 **Score:** 150 **Queries matched:** 4
autonomous glycyl radical cofactor GrcA [Klebsiella pneumoniae subsp. pneumoniae MGH 78578]
gi|157144488 **Mass:** 14376 **Score:** 150 **Queries matched:** 4
autonomous glycyl radical cofactor GrcA [Citrobacter koseri ATCC BAA-895]
gi|198244654 **Mass:** 14403 **Score:** 150 **Queries matched:** 4
autonomous glycyl radical cofactor [Salmonella enterica subsp. enterica serovar Dublin str. CT_02021853]
gi|237705089 **Mass:** 15033 **Score:** 150 **Queries matched:** 4
yfiD [Escherichia sp. 3_2_53FAA]
gi|237732565 **Mass:** 15009 **Score:** 150 **Queries matched:** 4
protein yfiD [Citrobacter sp. 30_2]
gi|261340878 **Mass:** 14490 **Score:** 150 **Queries matched:** 4
autonomous glycyl radical cofactor GrcA [Enterobacter cancerogenus ATCC 35316]
gi|262040312 **Mass:** 14378 **Score:** 150 **Queries matched:** 4
autonomous glycyl radical cofactor [Klebsiella pneumoniae subsp. rhinoscleromatis ATCC 13884]

Peptide matches not assigned to protein hits: (no details means no match)

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 96	681.3590	1360.7034	1360.6470	0.0564	2	49	0.36	1	KSGGKSPEDEAEK
☑ 95	679.3320	1356.6494	1356.7976	-0.1482	2	37	6.7	1	LNDIDSILKAKK
☑ 220	1113.0770	2224.1394	2224.0035	0.1360	2	36	5.9	1	MCAEVARRGGELVDGDGLMR + 2 Oxidation (M)
☑ 83	629.3340	1256.6534	1256.7816	-0.1281	2	33	15	1	IKASLQKTIQK
☑ 137	553.9480	1658.8222	1658.8410	-0.0188	2	32	16	1	TMARGLRELSDPEGK

Mascot Search Results

Protein View

Match to: **gi|261340878** Score: **150**
autonomous glycyl radical cofactor GrcA [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\meng\20080728\SSP4008_RB7_01_222.d\SSP4008_RB7_01_222.mgf

Nominal mass (M_r): **14490**; Calculated pI value: **4.85**
NCBI BLAST search of gi|261340878 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **29%**

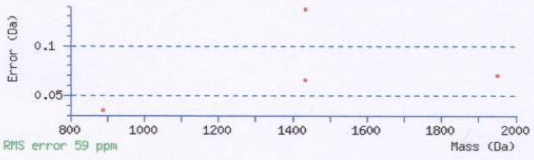
Matched peptides shown in **Bold Red**

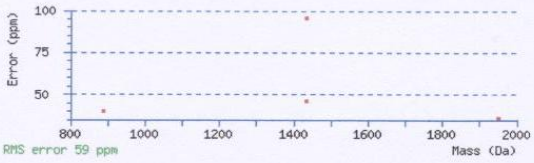
```
  1 MITGIQITKA ANDDLLNSFW LLDSEKNEAR CVVAKAGFAE DEIVPVNKLG
 51 EIEYREIPME VQPEVRVEGG QHLNVNVLRR ETLMDAVEHP EKYPQLTIRV
101 SGYAVRFNSL TPEQQRDVIA RTFTESL
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
10 - 26	976.0110	1950.0074	1949.9370	0.0704	0	K.AANDDLLNSFWLLDSEK.N (Ions score 75)
67 - 79	478.9540	1433.8402	1433.7739	0.0663	0	R.VEGGQHLNVNVLR.R (Ions score 37)
67 - 79	717.9630	1433.9114	1433.7739	0.1375	0	R.VEGGQHLNVNVLR.R (Ions score 39)
93 - 99	445.7760	889.5374	889.5021	0.0353	0	K.YPQLTIR.V (Ions score 38)






```
LOCUS       ZP_05968736              127 aa            linear   BCT 15-OCT-2009
DEFINITION  autonomous glycyl radical cofactor GrcA [Enterobacter cancerogenus
            ATCC 35316].
ACCESSION   ZP_05968736
VERSION     ZP_05968736.1  GI:261340878
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000010.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 127)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 127)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000010.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
```

```
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..127
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..127
                     /product="autonomous glycyl radical cofactor GrcA"
                     /calculated_mol_wt=14310
     Region          <69..123
                     /region_name="RNR_PFL"
                     /note="RNR_PFL. Ribonucleotide reductase  (RNR) and
                     pyruvate formate lyase (PFL) have a structurally similar
                     ten-stranded alpha-beta barrel active site domain and are
                     believed to have diverged from a common ancestor.  RNRs
                     are found in all organisms and...; cl09939"
                     /db_xref="CDD:142336"
     CDS             1..127
                     /locus_tag="EcanA3_020100010438"
                     /coded_by="complement(NZ_ABWM02000010.1:359251..359634)"
                     /note="COG3445 Acid-induced glycyl radical enzyme"
                     /transl_table=11
                     /db_xref="CDD:138424"
```

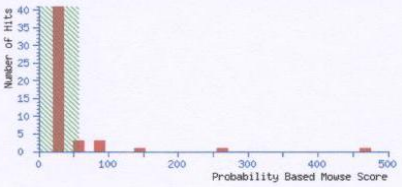
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 4602
MS data file    : D:\Data\Lakshmy\120808\260808\SSP 4602_RG21_01_403.d\SSP 4602_RG21_01_403.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 5 Jan 2010 at 11:19:04 GMT
Protein hits    : gi|261339284 outer membrane protein A [Enterobacter cancerogenus ATCC 35316]
                  gi|62901665  OmpA [Cronobacter sakazakii]
                  gi|157146352 outer membrane protein A [Citrobacter koseri ATCC BAA-895]
                  gi|262039986 outer membrane protein A [Klebsiella pneumoniae subsp. rhinoscleromatis ATCC 13884]
                  gi|14578722  major outer membrane protein [Pectobacterium cypripedii]
                  gi|129136    RecName: Full=Outer membrane protein A; Flags: Precursor
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|261339284 **Mass:** 37561 **Score:** 468 **Queries matched:** 16 **emPAI:** 0.53
outer membrane protein A [Enterobacter cancerogenus ATCC 35316]
☐ Check to include this hit in error tolerant search

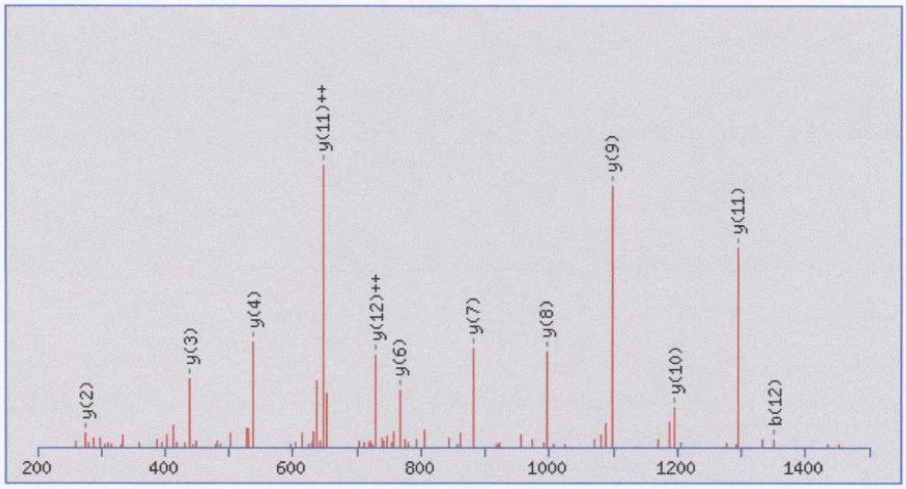
	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	84	542.2320	1082.4494	1082.5397	-0.0902	0	39	3.3	1	K.SDVLFNFNK.A
☑	94	604.8540	1207.6934	1207.6449	0.0486	0	44	1.1	1	R.AQSVVDYLVSK.G
☑	99	632.8000	1263.5854	1263.6460	-0.0605	0	65	0.01	1	K.DGSVVVLGFTDR.I
☑	104	455.2530	1362.7372	1363.7460	-1.0088	1	(50)	0.26	1	K.RAQSVVDYLVSK.G
☑	105	682.3840	1362.7534	1363.7460	-0.9925	1	51	0.25	1	K.RAQSVVDYLVSK.G
☑	127	813.6590	1625.3034	1625.7937	-0.4903	0	(22)	2.3e+02	1	K.LGYPVTDDLDVYTR.L
☑	128	813.8930	1625.7714	1625.7937	-0.0223	0	90	2.7e-05	1	K.LGYPVTDDLDVYTR.L
☑	129	813.9310	1625.8474	1625.7937	0.0537	0	(62)	0.016	1	K.LGYPVTDDLDVYTR.L
☑	130	813.9930	1625.9714	1625.7937	0.1777	0	(58)	0.039	1	K.LGYPVTDDLDVYTR.L
☑	131	814.4100	1626.8054	1625.7937	1.0117	0	(69)	0.0038	1	K.LGYPVTDDLDVYTR.L
☑	228	783.0930	2346.2572	2346.1856	0.0716	0	57	0.046	1	R.FGQQEDAPVVAPAPAPAPEVQTK.H
☑	243	890.8220	2669.4442	2670.3864	-0.9423	0	(33)	9.2	1	K.ATLKPEGQQALDQLYTQLSNLDPK.D
☑	244	891.1400	2670.3982	2670.3864	0.0117	0	52	0.14	1	K.ATLKPEGQQALDQLYTQLSNLDPK.D
	245	891.4750	2671.4032	2670.3864	1.0167	0	(21)	1.7e+02	8	K.ATLKPEGQQALDQLYTQLSNLDPK.D
☑	192	1017.1370	3048.3892	3048.3520	0.0372	0	71	0.003	1	R.ADSSNSIAGDDHDTGVSPVFAGGVEWAMTR.D
☑	195	1022.5150	3064.5232	3064.3469	0.1763	0	(24)	1.6e+02	1	R.ADSSNSIAGDDHDTGVSPVFAGGVEWAMTR.D + Oxidation (M)

2. gi|62901665 **Mass:** 37147 **Score:** 258 **Queries matched:** 9 **emPAI:** 0.29
OmpA [Cronobacter sakazakii]
☐ Check to include this hit in error tolerant search

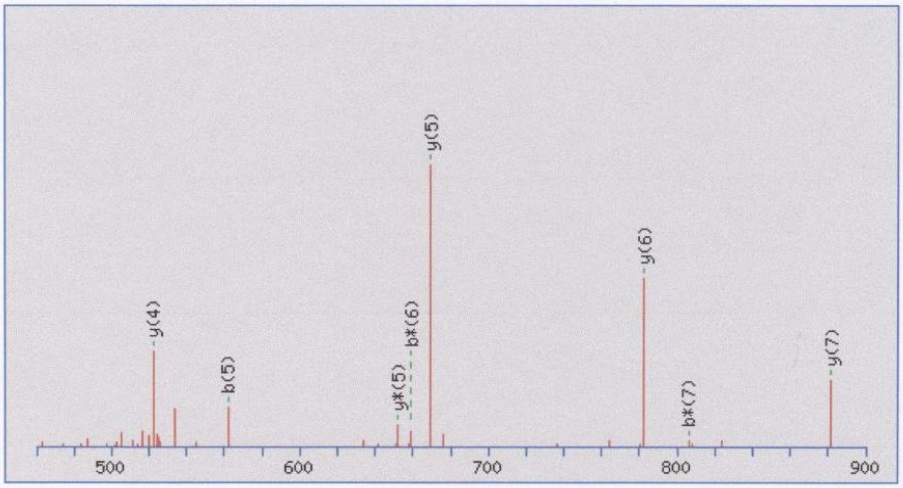
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
84	542.2320	1082.4494	1082.5397	-0.0902	0	39	3.3	1	K.SDVLFNFNK.A
99	632.8000	1263.5854	1263.6460	-0.0605	0	65	0.01	1	K.DGSVVVLGFTDR.I
127	813.6590	1625.3034	1625.7937	-0.4903	0	(22)	2.3e+02	1	K.LGYPVTDDLDVYTR.L
128	813.8930	1625.7714	1625.7937	-0.0223	0	90	2.7e-05	1	K.LGYPVTDDLDVYTR.L
129	813.9310	1625.8474	1625.7937	0.0537	0	(62)	0.016	1	K.LGYPVTDDLDVYTR.L
130	813.9930	1625.9714	1625.7937	0.1777	0	(58)	0.039	1	K.LGYPVTDDLDVYTR.L
131	814.4100	1626.8054	1625.7937	1.0117	0	(69)	0.0038	1	K.LGYPVTDDLDVYTR.L
192	1017.1370	3048.3892	3048.3520	0.0372	0	65	0.012	2	R.ADSSSNIAGDDHDTGVSPVFAGGVEWAMTR.D
195	1022.5150	3064.5232	3064.3469	0.1763	0	(24)	1.6e+02	1	R.ADSSSNIAGDDHDTGVSPVFAGGVEWAMTR.D + Oxidation (M)

Enterobacter cancerogenus SSP 4602 peptide sequences

LGYPVTDDLDVYTR – Ion score 90



SDVLFNFNK- Ion score 39



Mascot Search Results

Protein View

Match to: **gi|261339284** Score: **468**
outer membrane protein A [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\260808\SSP 4602_RG21_01_403.d\SSP 4602_RG21_01_403.mgf

Nominal mass (M_r): **37561**; Calculated pI value: **5.19**
NCBI BLAST search of gi|261339284 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **35%**

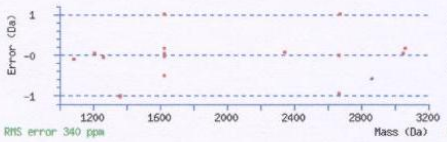
Matched peptides shown in **Bold Red**

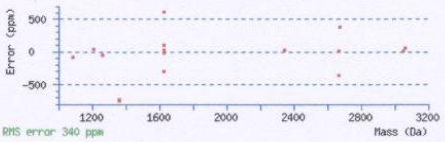
```
  1 MKKTAIAIAV ALAGFATVAQ AAPKDNTWYA GGKLGWSQFH DTGWYNSSLN
 51 NDGPTHESQL GAGAFGGYQV NPYVGFEMGY DWLGRMPYKG DNVNGAFKAQ
101 GVQLTAKLGY PVTDDLDVYT RLGGMVWRAD SSNSIAGDDH DTGVSPVFAG
151 GVEWAMTRDI ATRLEYQWVN NIGDGATVGV RPDNGMLSVG VSYRFGQQED
201 APVVAPAPAP APEVQTKHFT LKSDVLFNFN KATLKPEGQQ ALDQLYTQLS
251 NLDPKDGSVV VLGFTDRIGS DAYNQGLSEK RAQSVVDYLV SKGIPANKIS
301 PRGMGESNPV TGNTCDNVKP RAALIDCLAP DRRVEIEVKG IKDVVTQPAA
351
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

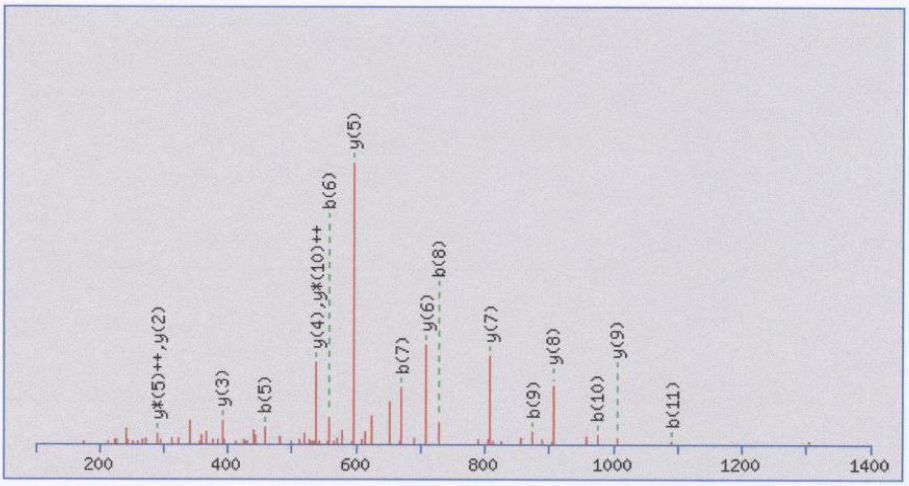
Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
108 - 121	813.6590	1625.3034	1625.7937	-0.4903	0	K.LGYPVTDDLDVYTR.L (Ions score 22)
108 - 121	813.8930	1625.7714	1625.7937	-0.0223	0	K.LGYPVTDDLDVYTR.L (Ions score 90)
108 - 121	813.9310	1625.8474	1625.7937	0.0537	0	K.LGYPVTDDLDVYTR.L (Ions score 62)
108 - 121	813.9930	1625.9714	1625.7937	0.1777	0	K.LGYPVTDDLDVYTR.L (Ions score 58)
108 - 121	814.4100	1626.8054	1625.7937	1.0117	0	K.LGYPVTDDLDVYTR.L (Ions score 69)
129 - 158	1017.1370	3048.3892	3048.3520	0.0372	0	R.ADSSNSIAGDDHDTGVSPVFAGGVEWAMTR.D (Ions score 71)
129 - 158	1022.5150	3064.5232	3064.3469	0.1763	0	R.ADSSNSIAGDDHDTGVSPVFAGGVEWAMTR.D Oxidation (M) (Ions score 24)
195 - 217	783.0930	2346.2572	2346.1856	0.0716	0	R.FGQQEDAPVVAPAPAPAPEVQTK.H (Ions score 57)
223 - 231	542.2320	1082.4494	1082.5397	-0.0902	0	K.SDVLFNFNK.A (Ions score 39)
232 - 255	890.8220	2669.4442	2670.3864	-0.9423	0	K.ATLKPEGQQALDQLYTQLSNLDPK.D (Ions score 33)
232 - 255	891.1400	2670.3982	2670.3864	0.0117	0	K.ATLKPEGQQALDQLYTQLSNLDPK.D (Ions score 52)
232 - 255	891.4750	2671.4032	2670.3864	1.0167	0	K.ATLKPEGQQALDQLYTQLSNLDPK.D (Ions score 21)
256 - 267	632.8000	1263.5854	1263.6460	-0.0605	0	K.DGSVVVLGFTDR.I (Ions score 65)
281 - 292	455.2530	1362.7372	1363.7460	-1.0088	1	K.RAQSVVDYLVSK.G (Ions score 50)
281 - 292	682.3840	1362.7534	1363.7460	-0.9925	1	K.RAQSVVDYLVSK.G (Ions score 51)
282 - 292	604.8540	1207.6934	1207.6449	0.0486	0	R.AQSVVDYLVSK.G (Ions score 44)






```
LOCUS       ZP_05967142              350 aa            linear   BCT 15-OCT-2009
DEFINITION  outer membrane protein A [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05967142
VERSION     ZP_05967142.1  GI:261339284
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000004.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 350)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 350)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000004.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
```

DGSVVVLGFTDR- Ion score 65



```
          manual curation.
          Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
          gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
          cancerogenus, not the sequenced strain), is a member of the
          Proteobacteria division of the domain bacteria and has been
          isolated from human feces. The sequenced strain was obtained from
          ATCC (ATCC 35316).

          We have performed one round of automated sequence improvement
          (pre-finishing), along with manual improvement that includes
          breaking apart any mis-assembly, and making manual joins where
          possible. Manual edits also are made where the consensus appears to
          be incorrect. All low quality data on the ends of contigs is
          removed. Contigs are ordered and oriented where possible.

          This work was supported by the National Human Genome Research
          Institute and the NIH Roadmap Human Microbiome Project
          [http://nihroadmap.nih.gov/hmp/].

          This is a reference genome for the Human Microbiome Project.
          This project is co-owned with the Human Microbiome Project DACC.
          Genome Coverage: 29.7x
          Sequencing Technology: 454.
          Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..350
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..350
                     /product="outer membrane protein A"
                     /calculated_mol_wt=37337
     Region          1..346
                     /region_name="PRK10808"
                     /note="outer membrane protein A; Reviewed; PRK10808"
                     /db_xref="CDD:138188"
     Region          23..196
                     /region_name="Surface_Ag_2"
                     /note="Surface antigen; c101155"
                     /db_xref="CDD:141128"
     Region          224..338
                     /region_name="OmpA_C-like"
                     /note="Peptidoglycan binding domains similar to the
                     C-terminal domain of outer-membrane protein OmpA; cd07185"
                     /db_xref="CDD:143586"
     Site            order(229..230,265..266,269,273..274,277,328,332)
                     /site_type="other"
                     /note="ligand binding site"
                     /db_xref="CDD:143586"
     CDS             1..350
                     /locus_tag="EcanA3_020100002425"
                     /coded_by="complement(NZ_ABWM02000004.1:545400..546452)"
                     /note="COG2885 Outer membrane protein and related
                     peptidoglycan-associated (lipo)proteins"
                     /transl_table=11
                     /db_xref="CDD:138188"
```

Mascot: http://www.matrixscience.com/

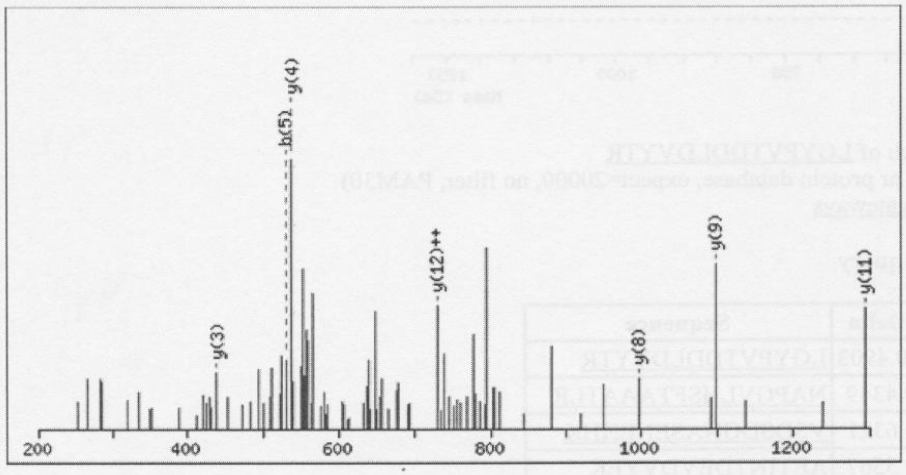
{MATRIX SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of **LGYPVTDDLDVYTR**
Found in **gi|261339284**, outer membrane protein A [Enterobacter cancerogenus ATCC 35316]

Match to Query 127: 1625.303448 from(813.659000,2+) intensity(2804959.0000)
Title: Cmpd 235, +MSn(814.24), 19.7 min
Data file D:\Data\Lakshmy\120808\260808\SSP 4602_RG21_01_403.d\SSP 4602_RG21_01_403.mgf

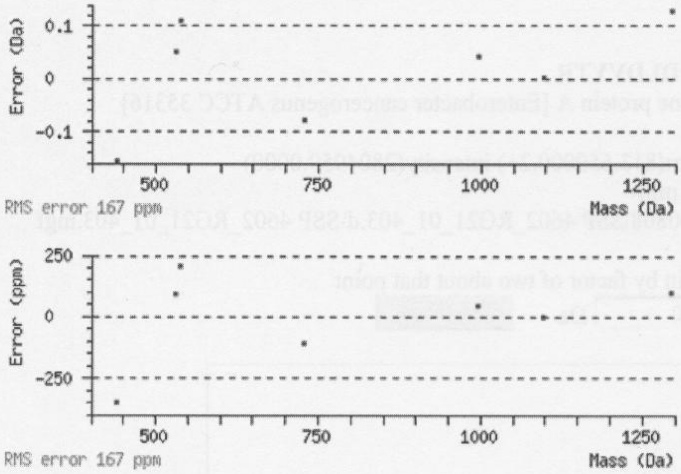
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 200 to 1300 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 1625.7937
Fixed modifications: Carboxymethyl (C)
Ions Score: 22 **Expect:** 2.3e+02
Matches (Bold Red): 7/104 fragment ions using 18 most intense peaks

#	a	a^{++}	b	b^{++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	86.0964	43.5519	114.0913	57.5493	L					14
2	143.1179	72.0626	171.1128	86.0600	G	1513.7169	757.3621	1496.6904	748.8488	13
3	306.1812	153.5942	334.1761	167.5917	Y	1456.6955	**728.8514**	1439.6689	720.3381	12
4	403.2340	202.1206	431.2289	216.1181	P	**1293.6321**	647.3197	1276.6056	638.8064	11
5	502.3024	251.6548	**530.2973**	265.6523	V	1196.5794	598.7933	1179.5528	590.2800	10
6	603.3501	302.1787	631.3450	316.1761	T	**1097.5109**	549.2591	1080.4844	540.7458	9
7	718.3770	359.6921	746.3719	373.6896	D	**996.4633**	498.7353	979.4367	490.2220	8
8	833.4040	417.2056	861.3989	431.2031	D	881.4363	441.2218	864.4098	432.7085	7
9	946.4880	473.7477	974.4829	487.7451	L	766.4094	383.7083	749.3828	375.1951	6
10	1061.5150	531.2611	1089.5099	545.2586	D	653.3253	327.1663	636.2988	318.6530	5
11	1160.5834	580.7953	1188.5783	594.7928	V	**538.2984**	269.6528	521.2718	261.1396	4
12	1323.6467	662.3270	1351.6416	676.3245	Y	**439.2300**	220.1186	422.2034	211.6053	3
13	1424.6944	712.8508	1452.6893	726.8483	T	276.1666	138.5870	259.1401	130.0737	2

14					R	175.1190	88.0631	158.0924	79.5498	1



NCBI **BLAST** search of LGYPVTDDLDVYTR
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
21.8	1625.7937	-0.4903	LGYPVTDDLDVYTR
11.7	1624.8685	0.4349	NAPGVLHSFTAAATLR
10.9	1624.6713	0.6321	VSDSDGNASENEHHK
10.6	1624.7468	0.5567	AETINTDEVDVYEK
10.6	1624.7986	0.5048	MRVKVILACTECK
10.6	1625.7607	-0.4573	FMTGLDDDLTQVGAK
9.6	1625.8049	-0.5015	EHGDLEGPLKAFADK
9.3	1625.9902	-0.6868	GIVSPILQVVKAMKK
9.1	1624.7072	0.5962	KENCPVASSSTMDAK
8.5	1625.7257	-0.4222	WFGSGGDGSVQGAMVR

Mascot: http://www.matrixscience.com/

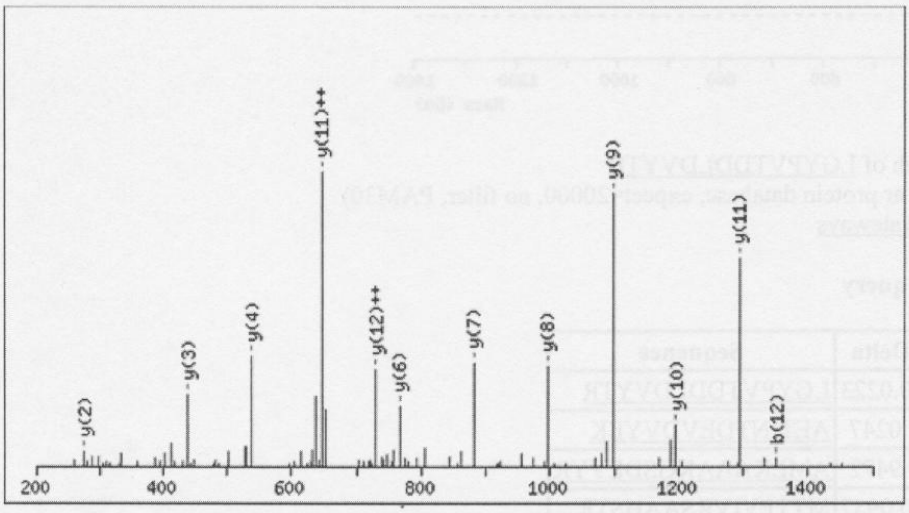
{MATRIX SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of **LGYPVTDDLDVYTR**
Found in **gi|261339284**, outer membrane protein A [Enterobacter cancerogenus ATCC 35316]

Match to Query 128: 1625.771448 from(813.893000,2+) intensity(85683688.0000)
Title: Cmpd 37, +MSn(814.35), 7.3 min
Data file D:\Data\Lakshmy\120808\260808\SSP 4602_RG21_01_403.d\SSP 4602_RG21_01_403.mgf

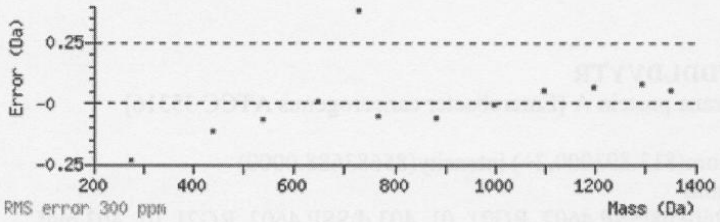
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 200 to 1500 Da Full range

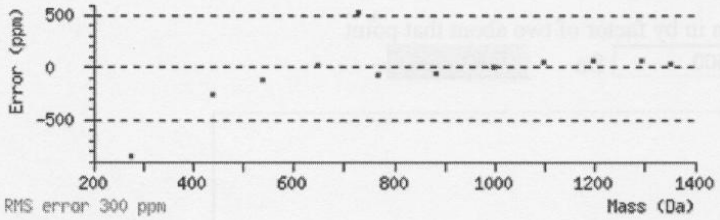


Monoisotopic mass of neutral peptide Mr(calc): 1625.7937
Fixed modifications: Carboxymethyl (C)
Ions Score: 90 **Expect:** 2.7e-05
Matches (Bold Red): 12/104 fragment ions using 15 most intense peaks

#	a	a^{++}	b	b^{++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	86.0964	43.5519	114.0913	57.5493	L					14
2	143.1179	72.0626	171.1128	86.0600	G	1513.7169	757.3621	1496.6904	748.8488	13
3	306.1812	153.5942	334.1761	167.5917	Y	1456.6955	**728.8514**	1439.6689	720.3381	12
4	403.2340	202.1206	431.2289	216.1181	P	**1293.6321**	**647.3197**	1276.6056	638.8064	11
5	502.3024	251.6548	530.2973	265.6523	V	**1196.5794**	598.7933	1179.5528	590.2800	10
6	603.3501	302.1787	631.3450	316.1761	T	**1097.5109**	549.2591	1080.4844	540.7458	9
7	718.3770	359.6921	746.3719	373.6896	D	**996.4633**	498.7353	979.4367	490.2220	8
8	833.4040	417.2056	861.3989	431.2031	D	**881.4363**	441.2218	864.4098	432.7085	7
9	946.4880	473.7477	974.4829	487.7451	L	**766.4094**	383.7083	749.3828	375.1951	6
10	1061.5150	531.2611	1089.5099	545.2586	D	653.3253	327.1663	636.2988	318.6530	5
11	1160.5834	580.7953	1188.5783	594.7928	V	**538.2984**	269.6528	521.2718	261.1396	4
12	1323.6467	662.3270	**1351.6416**	676.3245	Y	**439.2300**	220.1186	422.2034	211.6053	3
13	1424.6944	712.8508	1452.6893	726.8483	T	**276.1666**	138.5870	259.1401	130.0737	2

14					R	175.1190	88.0631	158.0924	79.5498	1





NCBI **BLAST** search of LGYPVTDDLDVYTR
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
90.0	1625.7937	-0.0223	LGYPVTDDLDVYTR
33.0	1624.7468	1.0247	AETINTDEVDVYEK
28.1	1624.8242	0.9472	AMLKAAAKGSDEVYR
23.7	1625.8672	-0.0957	MTTPVIVRSAAHSTR
23.2	1624.7733	0.9982	KFYDDKGNEVPGEK
20.3	1625.6748	0.0966	MECACARAHVYTR
19.9	1625.9002	-0.1287	RPGRVXFVDLSEVR
18.5	1625.6553	0.1161	DDHAEEPSDEAQQR
18.4	1625.7798	-0.0083	DLQRDTPFNTYTR
17.6	1625.8157	-0.0442	GMSYLAEMLSLNGIK

Mascot: http://www.matrixscience.com/

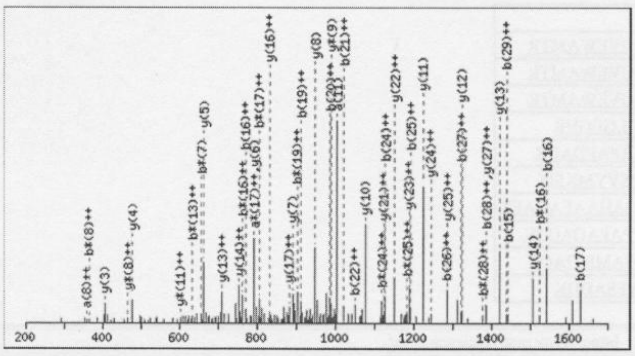
MATRIX SCIENCE Mascot Search Results

Peptide View

MS/MS Fragmentation of **ADSSNSIAGDDHDTGVSPVFAGGVEWAMTR**
Found in gi|**261339284**, outer membrane protein A [Enterobacter cancerogenus ATCC 35316]

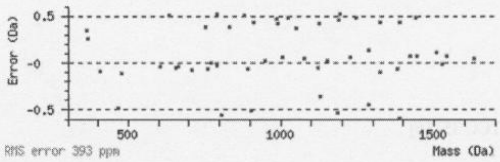
Match to Query 192: 3048.389172 from(1017.137000,3+) intensity(4976399.0000)
Title: Cmpd 86, +MSn(1017.75), 9.2 min
Data file D:\Data\Lakshmy\120808\260808\SSP 4602_RG21_01_403.d\SSP 4602_RG21_01_403.mgf

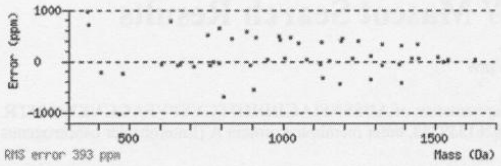
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 200 to 1700 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 3048.3520
Fixed modifications: Carboxymethyl (C)
Ions Score: 71 **Expect:** 0.003
Matches (Bold Red): 51/332 fragment ions using 88 most intense peaks

#	a	a^{++}	a*	a^{*++}	b	b^{++}	b*	b^{*++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	44.0495	22.5284			72.0444	36.5258			A					30
2	159.0764	80.0418			187.0713	94.0393			D	2978.3221	1489.6647	2961.2956	1481.1514	29
3	246.1084	123.5579			274.1034	137.5553			S	2863.2952	1432.1512	2846.2686	1423.6379	28
4	333.1405	167.0739			361.1354	181.0713			S	2776.2631	**1388.6352**	2759.2366	1380.1219	27
5	447.1834	224.0953	430.1569	215.5821	475.1783	238.0928	458.1518	229.5795	N	2689.2311	1345.1192	2672.2046	1336.6059	26
6	534.2154	267.6114	517.1889	259.0981	562.2103	281.6088	545.1838	273.0955	S	2575.1882	**1288.0977**	2558.1616	1279.5845	25
7	647.2995	324.1534	630.2729	315.6401	675.2944	338.1508	**658.2679**	329.6376	I	2488.1562	**1244.5817**	2471.1296	1236.0684	24
8	718.3366	**359.6719**	701.3101	351.1587	746.3315	373.6694	729.3050	**365.1561**	A	2375.0721	**1188.0397**	2358.0455	1179.5264	23
9	775.3581	388.1827	758.3315	379.6694	803.3530	402.1801	786.3264	393.6669	G	2304.0350	**1152.5211**	2287.0084	1144.0079	22
10	890.3850	445.6961	873.3585	437.1829	918.3799	459.6936	901.3534	451.1803	D	2247.0135	**1124.0104**	2229.9870	1115.4971	21
11	**1005.4120**	503.2096	988.3854	494.6963	1033.4069	517.2071	1016.3803	508.6938	D	2131.9866	1066.4969	2114.9600	1057.9836	20
12	1142.4709	571.7391	1125.4443	563.2258	1170.4658	585.7365	1153.4392	577.2233	H	2016.9596	1008.9835	1999.9331	1000.4702	19
13	1257.4978	629.2525	1240.4713	620.7393	1285.4927	643.2500	1268.4662	**634.7367**	D	1879.9007	940.4540	1862.8742	931.9407	18
14	1358.5455	679.7764	1341.5189	671.2631	1386.5404	693.7738	1369.5139	685.2606	T	1764.8738	**882.9405**	1747.8472	874.4272	17
15	1415.5670	708.2871	1398.5404	699.7738	**1443.5619**	722.2846	1426.5353	713.7713	G	1663.8261	**832.4167**	1646.7995	823.9034	16
16	1514.6354	757.8213	1497.6088	749.3080	**1542.6303**	**771.8188**	**1525.6037**	**763.3055**	V	1606.8046	803.9060	1589.7781	795.3927	15
17	1601.6674	801.3373	1584.6408	**792.8241**	**1629.6623**	815.3348	1612.6358	**806.8215**	S	**1507.7362**	**754.3717**	1490.7097	745.8585	14
18	1698.7202	849.8637	1681.6936	841.3504	1726.7151	863.8612	1709.6885	855.3479	P	**1420.7042**	**710.8557**	1403.6776	702.3425	13
19	1797.7886	899.3979	1780.7620	890.8847	1825.7835	**913.3954**	1808.7569	**904.8821**	V	**1323.6514**	662.3293	1306.6249	653.8161	12
20	1944.8570	972.9321	1927.8304	964.4189	1972.8519	**986.9296**	1955.8254	978.4163	F	**1224.5830**	612.7951	1207.5565	**604.2819**	11
21	2015.8941	1008.4507	1998.8676	999.9374	2043.8890	**1022.4481**	2026.8625	1013.9349	A	**1077.5146**	539.2609	1060.4880	530.7477	10
22	2072.9156	1036.9614	2055.8890	1028.4481	2100.9105	**1050.9589**	2083.8839	1042.4456	G	1006.4775	503.7424	**989.4509**	495.2291	9
23	2129.9370	1065.4722	2112.9105	1056.9589	2157.9319	1079.4696	2140.9054	1070.9563	G	**949.4560**	475.2316	932.4295	**466.7184**	8
24	2229.0054	1115.0064	2211.9789	1106.4931	2257.0004	**1129.0038**	2239.9738	**1120.4905**	V	**892.4346**	446.7209	875.4080	438.2076	7
25	2358.0480	1179.5277	2341.0215	1171.0144	2386.0430	**1193.5251**	2369.0164	**1185.0118**	E	**793.3661**	397.1867	776.3396	388.6734	6
26	2544.1274	1272.5673	2527.1008	1264.0540	2572.1223	**1286.5648**	2555.0957	1278.0515	W	**664.3235**	332.6654	647.2970	324.1521	5
27	2615.1645	1308.0859	2598.1379	1299.5726	2643.1594	**1322.0833**	2626.1328	1313.5701	A	**478.2442**	239.6258	461.2177	231.1125	4
28	2746.2050	1373.6061	2729.1784	1365.0928	2774.1999	**1387.6036**	2757.1733	**1379.0903**	M	**407.2071**	204.1072	390.1806	195.5939	3
29	2847.2526	1424.1300	2830.2261	1415.6167	2875.2475	**1438.1274**	2858.2210	1429.6141	T	276.1666	138.5870	259.1401	130.0737	2
30									R	175.1190	88.0631	158.0924	79.5498	1





NCBI **BLAST** search of ADSSNSIAGDDHDTGVSPVFAGGVEWAMTR
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
71.0	3048.3520	0.0372	ADSSNSIAGDDHDTGVSPVFAGGVEWAMTR
64.9	3048.3520	0.0372	ADSSSNIAGDDHDTGVSPVFAGGVEWAMTR
39.5	3047.3679	1.0212	ADSSNSIAGDNHDTGVSPVFAGGVEWAMTR
21.1	3048.4896	-0.1005	FIAVLTMSGIATAQSGTTSKFMQLGEGDR
20.4	3049.4662	-1.0771	MNLHAEDDVIAPVSTGTSVAALTPAPDADR
20.4	3048.5664	-0.1772	MLGPGSIVGNPLDAGFAAVVDPSVYMKSIK
18.9	3048.4860	-0.0968	NTNVPADSRDQAAALSAALTAGAHAAEAGAEK
18.3	3048.6001	-0.2109	MRLNHFACAAALLAATAAALVPAPAQAQTK
17.4	3047.4449	0.9442	VGGAQIDMMKYDMGGSAAVFGAMRTIAQR
17.4	3049.4530	-1.0638	QGSADATLSQPHYTLANARLSWFSADDK

Mascot: http://www.matrixscience.com/

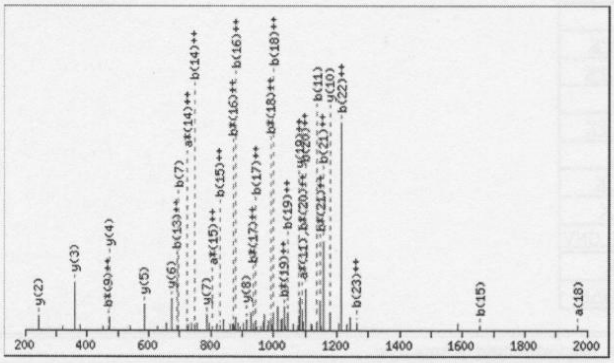
{MATRIX SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of **ATLKPEGQQALDQLYTQLSNLDPK**
Found in gi|**261339284**, outer membrane protein A [Enterobacter cancerogenus ATCC 35316]

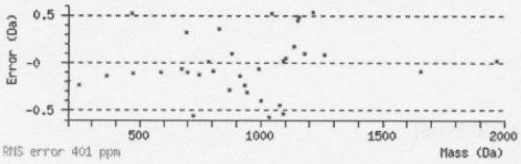
Match to Query 244: 2670.398172 from(891.140000,3+) intensity(46648868.0000)
Title: Cmpd 82, +MSn(891.82), 9.1 min
Data file D:\Data\Lakshmy\120808\260808\SSP 4602_RG21_01_403.d\SSP 4602_RG21_01_403.mgf

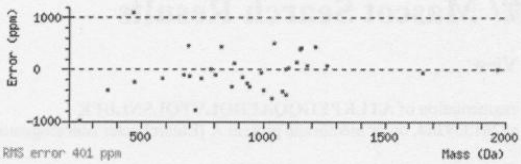
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 200 to 2000 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 2670.3864
Fixed modifications: Carboxymethyl (C)
Ions Score: 52 Expect: 0.14
Matches (Bold Red): 34/264 fragment ions using 62 most intense peaks

#	a	a^{++}	a*	a^{*++}	b	b^{++}	b*	b^{*++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	44.0495	22.5284			72.0444	36.5258			A					24
2	145.0972	73.0522			173.0921	87.0497			T	2600.3566	1300.6820	2583.3301	1292.1687	23
3	258.1812	129.5942			286.1761	143.5917			L	2499.3089	1250.1581	2482.2824	1241.6448	22
4	386.2762	193.6417	369.2496	185.1285	414.2711	207.6392	397.2445	199.1259	K	2386.2249	1193.6161	2369.1983	1185.1028	21
5	483.3289	242.1681	466.3024	233.6548	511.3239	256.1656	494.2973	247.6523	P	2258.1299	1129.5686	2241.1034	1121.0553	20
6	612.3715	306.6894	595.3450	298.1761	640.3665	320.6869	623.3399	312.1736	E	2161.0772	**1081.0422**	2144.0506	1072.5289	19
7	669.3930	335.2001	652.3665	326.6869	**697.3879**	349.1976	680.3614	340.6843	G	2032.0346	1016.5209	2015.0080	1008.0076	18
8	797.4516	399.2294	780.4250	390.7162	825.4465	413.2269	808.4199	404.7136	Q	1975.0131	988.0102	1957.9866	979.4969	17
9	925.5102	463.2587	908.4836	454.7454	953.5051	477.2562	936.4785	**468.7429**	Q	1846.9545	923.9809	1829.9280	915.4676	16
10	996.5473	498.7773	979.5207	490.2640	1024.5422	512.7747	1007.5156	504.2615	A	1718.8959	859.9516	1701.8694	851.4383	15
11	1109.6313	555.3193	**1092.6048**	546.8060	**1137.6263**	569.3168	1120.5997	560.8035	L	1647.8588	824.4331	1630.8323	815.9198	14
12	1224.6583	612.8328	1207.6317	604.3195	1252.6532	626.8302	1235.6266	618.3170	D	1534.7748	767.8910	1517.7482	759.3777	13
13	1352.7169	676.8621	1335.6903	668.3488	1380.7118	**690.8595**	1363.6852	682.3462	Q	1419.7478	710.3775	1402.7213	701.8643	12
14	1465.8009	733.4041	1448.7744	**724.8908**	1493.7958	**747.4016**	1476.7693	738.8883	L	1291.6892	646.3483	1274.6627	637.8350	11
15	1628.8643	814.9358	1611.8377	**806.4225**	**1656.8592**	**828.9332**	1639.8326	820.4199	Y	**1178.6052**	589.8062	1161.5786	581.2930	10
16	1729.9119	865.4596	1712.8854	856.9463	1757.9068	**879.4571**	1740.8803	**870.9438**	T	1015.5419	508.2746	998.5153	499.7613	9
17	1857.9705	929.4889	1840.9440	920.9756	1885.9654	**943.4863**	1868.9389	**934.9731**	Q	**914.4942**	457.7507	897.4676	449.2374	8
18	**1971.0546**	986.0309	1954.0280	977.5176	1999.0495	**1000.0284**	1982.0229	**991.5151**	L	**786.4356**	393.7214	769.4090	385.2082	7
19	2058.0866	1029.5469	2041.0600	1021.0337	2086.0815	**1043.5444**	2069.0550	**1035.0311**	S	**673.3515**	337.1794	656.3250	328.6661	6
20	2172.1295	1086.5684	2155.1030	1078.0551	2200.1244	**1100.5659**	2183.0979	**1092.0526**	N	**586.3195**	293.6634	569.2930	285.1501	5
21	2285.2136	1143.1104	2268.1870	1134.5972	2313.2085	**1157.1079**	2296.1820	**1148.5946**	L	**472.2766**	236.6419	455.2500	228.1287	4
22	2400.2405	1200.6239	2383.2140	1192.1106	2428.2354	**1214.6214**	2411.2089	1206.1081	D	**359.1925**	180.0999	342.1660	171.5866	3
23	2497.2933	1249.1503	2480.2667	1240.6370	2525.2882	**1263.1477**	2508.2617	1254.6345	P	**244.1656**	122.5864	227.1390	114.0731	2
24									K	147.1128	74.0600	130.0863	65.5468	1





NCBI **BLAST** search of ATLKPEGQQALDQLYTQLSNLDPK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
51.9	2670.3864	0.0117	ATLKPEGQQALDQLYTQLSNIDPK
51.9	2670.3864	0.0117	ATLKPEGQQALDQLYTQLSNLDPK
34.7	2669.4891	0.9091	IIFEIDQALGIENVTLDIKDGLLK
34.4	2671.4068	-1.0087	ATLKPEGQQALDQLYTQLTTLDPK
32.8	2671.2118	-0.8137	NREIHEQECFNLNPATNVMNPK
30.7	2670.2806	0.1175	DNQMRPLTDEELLQEQATALPDK
29.4	2669.4170	0.9812	QINDEMTVQNKALEIARQQLSLK
28.8	2670.3646	0.0335	AGNKGSEAAQGLLEMISLNQQILGNV
28.0	2671.2662	-0.8680	MVEPAEEQFSGFIFKIQANMDPK
27.9	2671.4908	-1.0927	FVIEEGIKESLLGGRTINATLVGGAK

Mascot: http://www.matrixscience.com/

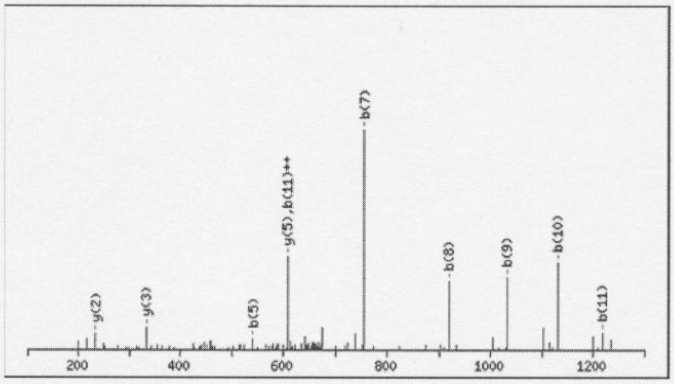
Mascot Search Results

Peptide View

MS/MS Fragmentation of **RAQSVVDYLVSK**
Found in **gi|261339284**, outer membrane protein A [Enterobacter cancerogenus ATCC 35316]

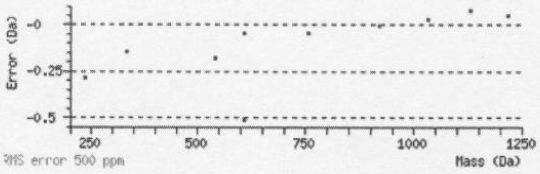
Match to Query 105: 1362.753448 from(682.384000,2+) intensity(2643055.0000)
Title: Cmpd 2, +MSn(683.26), 5.4 min
Data file D:\Data\Lakshmy\120808\260808\SSP 4602_RG21_01_403.d\SSP 4602_RG21_01_403.mgf

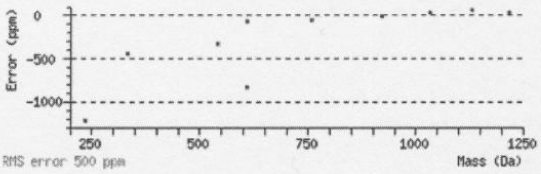
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 100 to 1300 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 1363.7460
Fixed modifications: Carboxymethyl (C)
Ions Score: 51 **Expect:** 0.25
Matches (Bold Red): 10/132 fragment ions using 11 most intense peaks

#	a	a^{++}	a*	a^{*++}	b	b^{++}	b*	b^{*++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	129.1135	65.0604	112.0869	56.5471	157.1084	79.0578	140.0818	70.5446	R					12
2	200.1506	100.5789	183.1240	92.0657	228.1455	114.5764	211.1190	106.0631	A	1208.6521	604.8297	1191.6256	596.3164	11
3	328.2092	164.6082	311.1826	156.0949	356.2041	178.6057	339.1775	170.0924	Q	1137.6150	569.3111	1120.5885	560.7979	10
4	415.2412	208.1242	398.2146	199.6110	443.2361	222.1217	426.2096	213.6084	S	1009.5564	505.2819	992.5299	496.7686	9
5	514.3096	257.6584	497.2831	249.1452	**542.3045**	271.6559	525.2780	263.1426	V	922.5244	461.7658	905.4979	453.2526	8
6	613.3780	307.1926	596.3515	298.6794	641.3729	321.1901	624.3464	312.6768	V	823.4560	412.2316	806.4294	403.7184	7
7	728.4050	364.7061	711.3784	356.1928	**756.3999**	378.7036	739.3733	370.1903	D	724.3876	362.6974	707.3610	354.1842	6
8	891.4683	446.2378	874.4417	437.7245	**919.4632**	460.2352	902.4367	451.7220	Y	**609.3606**	305.1840	592.3341	296.6707	5
9	1004.5524	502.7798	987.5258	494.2665	**1032.5473**	516.7773	1015.5207	508.2640	L	446.2973	223.6523	429.2708	215.1390	4
10	1103.6208	552.3140	1086.5942	543.8007	**1131.6157**	566.3115	1114.5891	557.7982	V	**333.2132**	167.1103	316.1867	158.5970	3
11	1190.6528	595.8300	1173.6262	587.3168	**1218.6477**	**609.8275**	1201.6212	601.3142	S	**234.1448**	117.5761	217.1183	109.0628	2
12									K	147.1128	74.0600	130.0863	65.5468	1





NCBI **BLAST** search of RAQSVVDYLVSK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
50.6	1363.7460	-0.9925	RAQSVVDYLVSK
36.3	1363.7248	-0.9714	HFSKANAYLVSK

36.3	1363.7711	-1.0176	LSEGQKIYIVSK
34.4	1363.6983	-0.9449	EAGLSEFANIVSK
28.9	1363.7347	-0.9812	EGKAKEIYIVASG
28.9	1363.6983	-0.9449	EGKAQEIYIVASG
28.9	1363.6983	-0.9449	EGKAQEIYIVXSG
27.8	1363.7096	-0.9561	FNAIVTRDDVSK
26.6	1363.6619	-0.9085	EGELQAQFEVSK
26.2	1363.8075	-1.0541	KILITGGTTFVSK

Mascot: http://www.matrixscience.com/

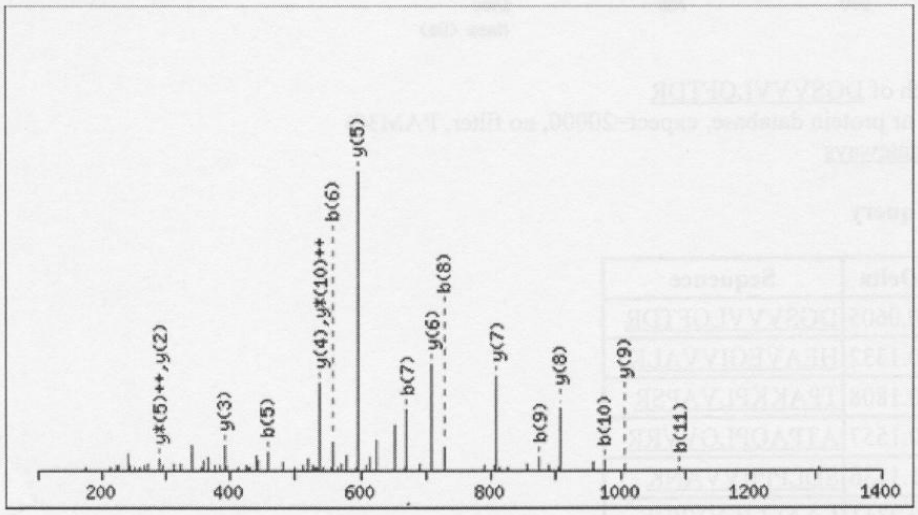
{MATRIX SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of **DGSVVVLGFTDR**
Found in **gi|261339284**, outer membrane protein A [Enterobacter cancerogenus ATCC 35316]

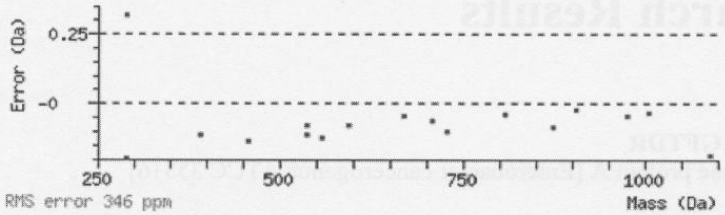
Match to Query 99: 1263.585448 from(632.800000,2+) intensity(34117664.0000)
Title: Cmpd 23, +MSn(633.16), 6.6 min
Data file D:\Data\Lakshmy\120808\260808\SSP 4602_RG21_01_403.d\SSP 4602_RG21_01_403.mgf

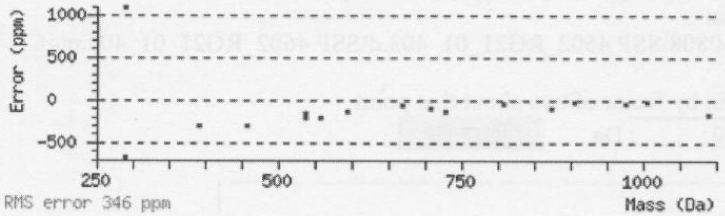
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 100 to 1400 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 1263.6460
Fixed modifications: Carboxymethyl (C)
Ions Score: 65 **Expect:** 0.01
Matches (Bold Red): 17/88 fragment ions using 23 most intense peaks

#	a	a^{++}	b	b^{++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	88.0393	44.5233	116.0342	58.5207	D					12
2	145.0608	73.0340	173.0557	87.0315	G	1149.6263	575.3168	1132.5997	566.8035	11
3	232.0928	116.5500	260.0877	130.5475	S	1092.6048	546.8060	1075.5782	**538.2928**	10
4	331.1612	166.0842	359.1561	180.0817	V	**1005.5728**	503.2900	988.5462	494.7767	9
5	430.2296	215.6184	**458.2245**	229.6159	V	**906.5043**	453.7558	889.4778	445.2425	8
6	529.2980	265.1527	**557.2930**	279.1501	V	**807.4359**	404.2216	790.4094	395.7083	7
7	642.3821	321.6947	**670.3770**	335.6921	L	**708.3675**	354.6874	691.3410	346.1741	6
8	699.4036	350.2054	**727.3985**	364.2029	G	**595.2835**	298.1454	578.2569	**289.6321**	5
9	846.4720	423.7396	**874.4669**	437.7371	F	**538.2620**	269.6346	521.2354	261.1214	4
10	947.5197	474.2635	**975.5146**	488.2609	T	**391.1936**	196.1004	374.1670	187.5872	3
11	1062.5466	531.7769	**1090.5415**	545.7744	D	**290.1459**	145.5766	273.1193	137.0633	2
12					R	175.1190	88.0631	158.0924	79.5498	1





NCBI **BLAST** search of DGSVVVLGFTDR
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
64.5	1263.6460	-0.0605	DGSVVVLGFTDR
43.3	1263.7187	-0.1332	HEAVEGIVVALK
35.8	1263.7663	-0.1808	TPAKKPLVAPSR
34.0	1263.7412	-0.1557	ATPAQPLGVVRR
33.2	1264.7391	-1.1536	SDLPPIIVVANK
31.6	1263.8166	-0.2311	LAALLILVVPDK
29.7	1263.8642	-0.2788	VVGLLILVVIAR
29.2	1263.8027	-0.2172	KIPQKIPTLAR
29.0	1262.6143	0.9712	GDIVNGIGFDEK
28.8	1263.7299	-0.1445	GLAAPKGLPEGVR

Mascot: http://www.matrixscience.com/

21.4	1208.6877	-0.9943	TVSVVEAVHIR
19.9	1208.6513	-0.9579	DQDQHKVVIK
19.8	1207.6019	0.0915	ASQVVQFCLR
19.7	1206.5993	1.0942	REVVEQEYR
19.2	1206.6179	1.0755	REVMPVYASR
19.1	1206.6608	1.0326	AQSVVNYLVSK
18.9	1208.5707	-0.8772	SISVVSNDMNK

Mascot: http://www.matrixscience.com/

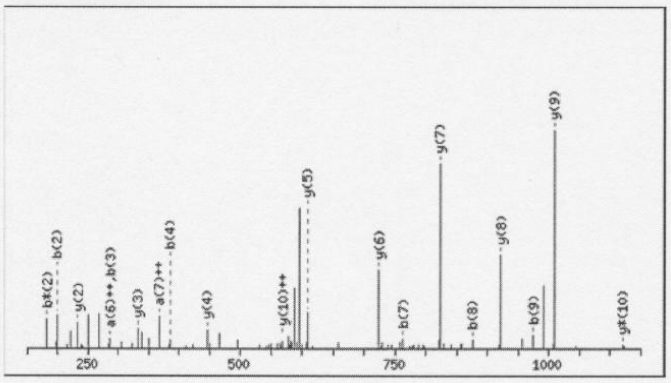
{MATRIX SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of **AQSVVDYLVSK**
Found in **gi|261339284**, outer membrane protein A [Enterobacter cancerogenus ATCC 35316]

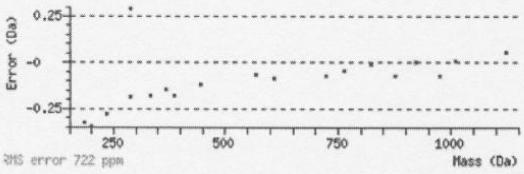
Match to Query 94: 1207.693448 from(604.854000,2+) intensity(17956968.0000)
Title: Cmpd 7, +MSn(605.23), 5.7 min
Data file D:\Data\Lakshmy\120808\260808\SSP 4602_RG21_01_403.d\SSP 4602_RG21_01_403.mgf

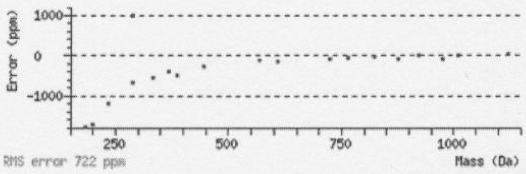
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 150 to 1150 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 1207.6449
Fixed modifications: Carboxymethyl (C)
Ions Score: 44 **Expect:** 1.1
Matches (Bold Red): 19/116 fragment ions using 47 most intense peaks

#	a	a^{++}	a*	a^{*++}	b	b^{++}	b*	b^{*++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	44.0495	22.5284			72.0444	36.5258			A					11
2	172.1081	86.5577	155.0815	78.0444	**200.1030**	100.5551	**183.0764**	92.0418	Q	1137.6150	**569.3111**	**1120.5885**	560.7979	10
3	259.1401	130.0737	242.1135	121.5604	**287.1350**	144.0711	270.1084	135.5579	S	**1009.5564**	505.2819	992.5299	496.7686	9
4	358.2085	179.6079	341.1819	171.0946	**386.2034**	193.6053	369.1769	185.0921	V	**922.5244**	461.7658	905.4979	453.2526	8
5	457.2769	229.1421	440.2504	220.6288	485.2718	243.1396	468.2453	234.6263	V	**823.4560**	412.2316	806.4294	403.7184	7
6	572.3039	**286.6556**	555.2773	278.1423	600.2988	300.6530	583.2722	292.1397	D	**724.3876**	362.6974	707.3610	354.1842	6
7	735.3672	**368.1872**	718.3406	359.6740	**763.3621**	382.1847	746.3355	373.6714	Y	**609.3606**	305.1840	592.3341	296.6707	5
8	848.4512	424.7293	831.4247	416.2160	**876.4462**	438.7267	859.4196	430.2134	L	**446.2973**	223.6523	429.2708	215.1390	4
9	947.5197	474.2635	930.4931	465.7502	**975.5146**	488.2609	958.4880	479.7477	V	**333.2132**	167.1103	316.1867	158.5970	3
10	1034.5517	517.7795	1017.5251	509.2662	1062.5466	531.7769	1045.5201	523.2637	S	**234.1448**	117.5761	217.1183	109.0628	2
11									K	147.1128	74.0600	130.0863	65.5468	1





NCBI **BLAST** search of AQSVVDYLVSK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
44.0	1207.6449	0.0486	AQSVVDYLVSK
29.4	1206.7085	0.9850	VVSVVQAGVPPR
24.0	1207.6673	0.0261	GHGQVIGNLVSK

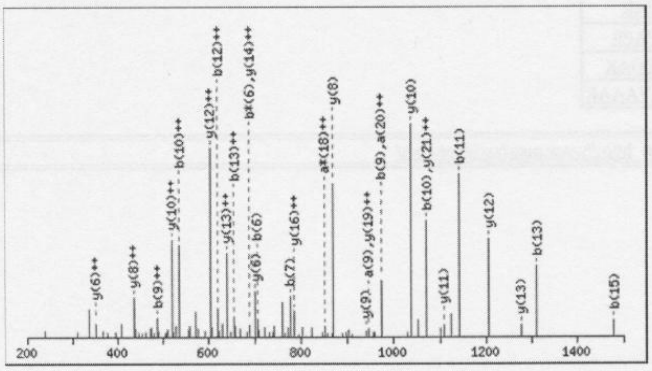
{MATRIX SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of **FGQQEDAPVVAPAPAPAPEVQTK**
Found in gi|**261339284**, outer membrane protein A [Enterobacter cancerogenus ATCC 35316]

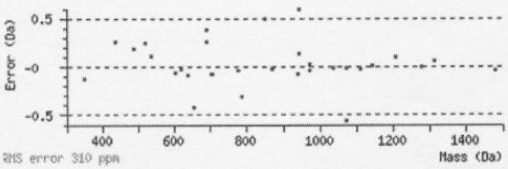
Match to Query 228: 2346.257172 from(783.093000,3+) intensity(68730400.0000)
Title: Cmpd 39, +MSn(783.69), 7.3 min
Data file D:\Data\Lakshmy\120808\260808\SSP 4602_RG21_01_403.d\SSP 4602_RG21_01_403.mgf

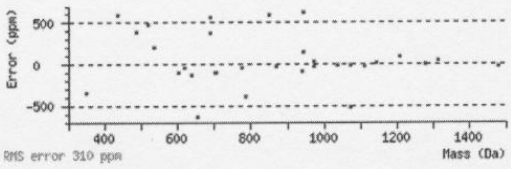
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 200 to 1500 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 2346.1856
Fixed modifications: Carboxymethyl (C)
Ions Score: 57 **Expect:** 0.046
Matches (Bold Red): 31/256 fragment ions using 40 most intense peaks

#	a	a^{++}	a*	a^{*++}	b	b^{++}	b*	b^{*++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	120.0808	60.5440			148.0757	74.5415			F					23
2	177.1022	89.0548			205.0972	103.0522			G	2200.1244	1100.5659	2183.0979	1092.0526	22
3	305.1608	153.0840	288.1343	144.5708	333.1557	167.0815	316.1292	158.5682	Q	2143.1030	**1072.0551**	2126.0764	1063.5419	21
4	433.2194	217.1133	416.1928	208.6001	461.2143	231.1108	444.1878	222.5975	Q	2015.0444	1008.0258	1998.0179	999.5126	20
5	562.2620	281.6346	545.2354	273.1214	590.2569	295.6321	573.2304	287.1188	E	1886.9858	**943.9965**	1869.9593	935.4833	19
6	677.2889	339.1481	660.2624	330.6348	**705.2838**	353.1456	**688.2573**	344.6323	D	1757.9432	879.4753	1740.9167	870.9620	18
7	748.3260	374.6667	731.2995	366.1534	**776.3210**	388.6641	759.2944	380.1508	A	1642.9163	821.9618	1625.8897	813.4485	17
8	845.3788	423.1930	828.3523	414.6798	873.3737	437.1905	856.3472	428.6772	P	1571.8792	**786.4432**	1554.8526	777.9299	16
9	**944.4472**	472.7272	927.4207	464.2140	**972.4421**	**486.7247**	955.4156	478.2114	V	1474.8264	737.9168	1457.7999	729.4036	15
10	1043.5156	522.2615	1026.4891	513.7482	**1071.5106**	**536.2589**	1054.4840	527.7456	V	1375.7580	**688.3826**	1358.7314	679.8694	14
11	1114.5528	557.7800	1097.5262	549.2667	**1142.5477**	571.7775	1125.5211	563.2642	A	**1276.6896**	**638.8484**	1259.6630	630.3352	13
12	1211.6055	606.3064	1194.5790	597.7931	1239.6004	**620.3039**	1222.5739	611.7906	P	**1205.6525**	**603.3299**	1188.6259	594.8166	12
13	1282.6426	641.8250	1265.6161	633.3117	**1310.6375**	**655.8224**	1293.6110	647.3091	A	**1108.5997**	554.8035	1091.5732	546.2902	11
14	1379.6954	690.3513	1362.6688	681.8381	1407.6903	704.3488	1390.6638	695.8355	P	**1037.5626**	**519.2849**	1020.5360	510.7717	10
15	1450.7325	725.8699	1433.7060	717.3566	**1478.7274**	739.8673	1461.7009	731.3541	A	**940.5098**	470.7586	923.4833	462.2453	9
16	1547.7853	774.3963	1530.7587	765.8830	1575.7802	788.3937	1558.7536	779.8805	P	**869.4727**	**435.2400**	852.4462	426.7267	8
17	1618.8224	809.9148	1601.7958	801.4016	1646.8173	823.9123	1629.7908	815.3990	A	772.4199	386.7136	755.3934	378.2003	7
18	1715.8751	858.4412	1698.8486	**849.9279**	1743.8701	872.4387	1726.8435	863.9254	P	**701.3828**	**351.1951**	684.3563	342.6818	6
19	1844.9177	922.9625	1827.8912	914.4492	1872.9127	936.9600	1855.8861	928.4467	E	604.3301	302.6687	587.3035	294.1554	5
20	1943.9862	972.4967	1926.9596	963.9834	1971.9811	986.4942	1954.9545	977.9809	V	475.2875	238.1474	458.2609	229.6341	4
21	2072.0447	1036.5260	2055.0182	1028.0127	2100.0396	1050.5235	2083.0131	1042.0102	Q	376.2191	188.6132	359.1925	180.0999	3
22	2173.0924	1087.0498	2156.0659	1078.5366	2201.0873	1101.0473	2184.0608	1092.5340	T	248.1605	124.5839	231.1339	116.0706	2
23									K	147.1128	74.0600	130.0863	65.5468	1





NCBI **BLAST** search of FGQQEDAPVVAPAPAPAPEVQTK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
56.9	2346.1856	0.0716	FGQQEDAPVVAPAPAPAPEVQTK
35.8	2346.1855	0.0716	FGQQEEAAPIAPAPAPAPEVQTK
23.3	2346.1420	0.1152	MDEDERMARAALSGLEANLVR
22.7	2346.2042	0.0530	VLMLVSGGLGQALWYSHDEKK
21.8	2345.2267	1.0304	FGQDDAVAPVVAPAPAPAPVVETK
20.1	2346.2154	0.0417	QVYGTVKGKLMAWDPVAQEAR
19.8	2345.1838	1.0734	EQGIDTVWFGAAAPLGLMAGVAR
19.7	2345.3008	0.9564	FNAPKAAPAPAPVPVAAPAPAPAPR
19.5	2347.2318	-0.9746	VAIISGAAQGMGAATARLFAAEGAK
18.3	2346.2880	-0.0308	SARADLLGPVGAPASGRAAAAVAAAR

Mascot: http://www.matrixscience.com/

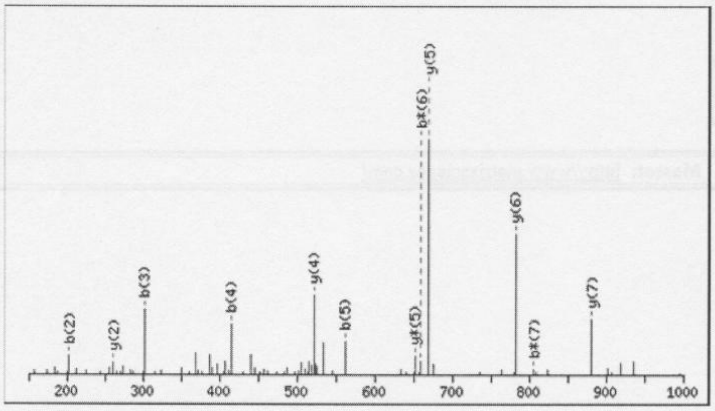
{MATRIX SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of **SDVLFNFNK**
Found in gi|**261339284**, outer membrane protein A [Enterobacter cancerogenus ATCC 35316]

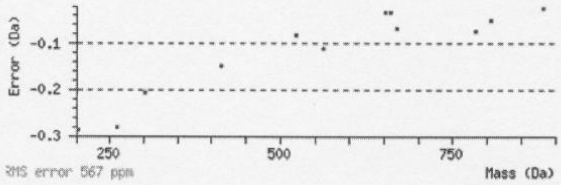
Match to Query 84: 1082.449448 from(542.232000,2+) intensity(14554407.0000)
Title: Cmpd 14, +MSn(543.11), 6.2 min
Data file D:\Data\Lakshmy\120808\260808\SSP 4602_RG21_01_403.d\SSP 4602_RG21_01_403.mgf

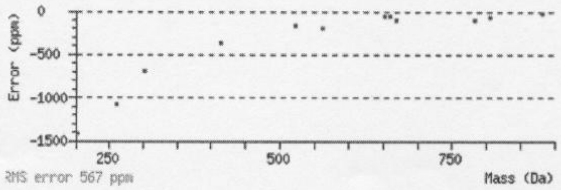
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 150 to 1000 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 1082.5397
Fixed modifications: Carboxymethyl (C)
Ions Score: 39 **Expect:** 3.3
Matches (Bold Red): 12/76 fragment ions using 17 most intense peaks

#	a	a^{++}	a*	a^{*++}	b	b^{++}	b*	b^{*++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	60.0444	30.5258			88.0393	44.5233			S					9
2	175.0713	88.0393			**203.0662**	102.0368			D	996.5149	498.7611	979.4884	490.2478	8
3	274.1397	137.5735			**302.1347**	151.5710			V	**881.4880**	441.2476	864.4614	432.7343	7
4	387.2238	194.1155			**415.2187**	208.1130			L	**782.4196**	391.7134	765.3930	383.2001	6
5	534.2922	267.6498			**562.2871**	281.6472			F	**669.3355**	335.1714	**652.3089**	326.6581	5
6	648.3352	324.6712	631.3086	316.1579	676.3301	338.6687	**659.3035**	330.1554	N	**522.2671**	261.6372	505.2405	253.1239	4
7	795.4036	398.2054	778.3770	389.6921	823.3985	412.2029	**806.3719**	403.6896	F	408.2241	204.6157	391.1976	196.1024	3
8	909.4465	455.2269	892.4199	446.7136	937.4414	469.2243	920.4149	460.7111	N	**261.1557**	131.0815	244.1292	122.5682	2
9									K	147.1128	74.0600	130.0863	65.5468	1





NCBI **BLAST** search of SDVLFNFNK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
38.9	1082.5397	-0.0902	SDVLFNFNK
37.4	1082.5318	-0.0824	SDVIFDMIK
37.4	1082.5318	-0.0824	SDVIFLMDK
32.4	1082.5396	-0.0902	AEELFGYVR
30.3	1082.6448	-0.1953	LVALQDGLVR
29.7	1082.5397	-0.0902	DEGLFTFVR
24.5	1082.6059	-0.1565	MFRVFIVR
24.5	1082.5253	-0.0758	MFTPGMIVR
23.8	1082.5760	-0.1266	FLSSDFIVR
22.9	1082.5720	-0.1226	ADPLDGQIVR

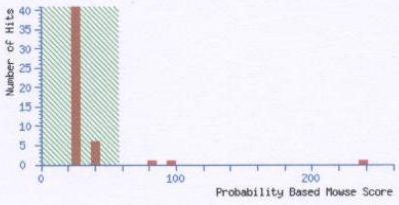
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 4606
MS data file    : D:\Data\Lakshmy\120808\241108\SSP 4606_RB21_01_773.d\SSP 4606_RB21_01_773.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 5 Jan 2010 at 11:20:40 GMT
Protein hits    : gi|50582463  oxidoreductase [Exiguobacterium acetylicum]
                  gi|146275769 short-chain dehydrogenase/reductase SDR [Novosphingobium aromaticivorans DSM 12444]
                  gi|148368    outer membrane protein II [Enterobacter aerogenes]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|50582463 **Mass:** 27360 **Score:** 238 **Queries matched:** 5 **emPAI:** 0.41
oxidoreductase [Exiguobacterium acetylicum]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 125	497.5760	1489.7062	1489.7413	-0.0351	0	53	0.15	1	K.VLAPAATETEFADR.S
☑ 155	573.2590	1716.7552	1716.8570	-0.1019	0	(25)	87	1	K.YTVITGASSGIGYETAK.L
☑ 156	859.3900	1716.7654	1716.8570	-0.0916	0	91	2.3e-05	1	K.YTVITGASSGIGYETAK.L
☑ 197	676.9390	2027.7952	2028.9752	-1.1801	0	(37)	3.6	1	K.SVDLADNQNVHDLYEGLK.E
☑ 198	1014.9250	2027.8354	2028.9752	-1.1398	0	94	7.5e-06	1	K.SVDLADNQNVHDLYEGLK.E

2. gi|146275769 **Mass:** 27018 **Score:** 97 **Queries matched:** 2 **emPAI:** 0.26
short-chain dehydrogenase/reductase SDR [Novosphingobium aromaticivorans DSM 12444]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 107	455.1670	1362.4792	1362.6204	-0.1413	0	52	0.14	1	K.FEDTPLSDFHR.V
☑ 211	700.5920	2098.7542	2099.0027	-0.2486	0	44	0.63	1	R.VNSVHPGGIDTPMLGSIMDK.Y + 2 Oxidation (M)

3. gi|148368 **Mass:** 25654 **Score:** 84 **Queries matched:** 2 **emPAI:** 0.13
outer membrane protein II [Enterobacter aerogenes]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 141	813.8050	1625.5954	1625.7937	-0.1983	0	48	0.32	1	K.LGYPVTDDLDVYTR.L
☑ 228	783.0500	2346.1282	2346.1856	-0.0574	0	36	5.2	1	R.FGQQEDAPVVAPAPAPAPEVQTK.H

Proteins matching the same set of peptides:
gi|261339284 **Mass:** 37561 **Score:** 84 **Queries matched:** 2
outer membrane protein A [Enterobacter cancerogenus ATCC 35316]

Peptide matches not assigned to protein hits: (no details means no match)

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 101	651.8070	1301.5994	1302.5914	-0.9920	0	40	2.9	1	GYYCSLLAEAR
☑ 108	692.3100	1382.6054	1382.8245	-0.2191	2	39	3.1	1	LSIALGKGSKNPAK
☑ 120	738.3480	1474.6814	1475.8209	-1.1394	1	36	7.1	1	TRTTFVQQLINR
☑ 229	789.7300	2366.1682	2366.3533	-0.1851	2	34	9.9	1	GDNGVGKTTLLKILAGLLQPSSGK

Mascot Search Results

Protein View

Match to: **gi|261339284** Score: **84**
outer membrane protein A [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\241108\SSP 4606_RB21_01_773.d\SSP 4606_RB21_01_773.mgf

Nominal mass (M_r): **37561**; Calculated pI value: **5.19**
NCBI BLAST search of gi|261339284 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **10%**

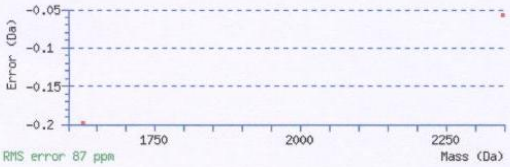
Matched peptides shown in **Bold Red**

```
  1 MKKTAIAIAV ALAGFATVAQ AAPKDNTWYA GGKLGWSQFH DTGWYNSSLN
 51 NDGPTHESQL GAGAFGGYQV NPYVGFEMGY DWLGRMPYKG DNVNGAFKAQ
101 GVQLTAKLGY PVTDDLDVYT RLGGMVWRAD SSNSIAGDDH DTGVSPVFAG
151 GVEWAMTRDI ATRLEYQWVN NIGDGATVGV RPDNGMLSVG VSYRFGQQED
201 APVVAPAPAP APEVQTKHFT LKSDVLFNFN KATLKPEGQQ ALDQLYTQLS
251 NLDPKDGSVV VLGFTDRIGS DAYNQGLSEK RAQSVVDYLV SKGIPANKIS
301 PRGMGESNPV TGNTCDNVKP RAALIDCLAP DRRVEIEVKG IKDVVTQPAA
351
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
108 - 121	813.8050	1625.5954	1625.7937	-0.1983	0	K.LGYPVTDDLDVYTR.L (Ions score 48)
195 - 217	783.0500	2346.1282	2346.1856	-0.0574	0	R.FGQQEDAPVVAPAPAPAPEVQTK.H (Ions score 36)



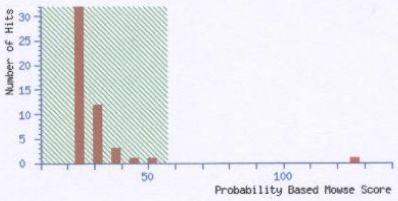

```
LOCUS       ZP_05967142              350 aa            linear   BCT 15-OCT-2009
DEFINITION  outer membrane protein A [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05967142
VERSION     ZP_05967142.1  GI:261339284
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000004.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 350)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 350)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000004.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
```

Mascot Search Results

User : Lakshmy Manickan
Email : lakshmy.manickan@unn.ac.uk
Search title : SSP 5003
MS data file : D:\Data\Lakshmy\120808\280808\SSP 5003_RJ6_01_513.d\SSP 5003_RJ6_01_513.mgf
Database : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy : Bacteria (Eubacteria) (5690016 sequences)
Timestamp : 5 Jan 2010 at 11:22:03 GMT
Protein hits : gi|261342318 hypothetical protein EcanA3_17757 [Enterobacter cancerogenus ATCC 35316]

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 Max. number of hits AUTO
Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0
Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned Decreasing Score Require bold red ☐

Select All | Select None | Search Selected | ☐ Error tolerant

1. gi|261342318 **Mass:** 26485 **Score:** 126 **Queries matched:** 2 **emPAI:** 0.27
hypothetical protein EcanA3_17757 [Enterobacter cancerogenus ATCC 35316]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	112	556.2900	1665.8482	1665.8726	-0.0244	0	36	7.1	1	R.VAQYLSFLEQNGVAK.K
☑	166	628.6650	1882.9732	1882.9537	0.0195	0	90	2.4e-05	1	K.AAIDDAIHQAQQLASGFK.G

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	67	679.3770	1356.7394	1356.7976	-0.0582	2	48	0.45	1	LNDIDSILKAKK
☑	233	749.4760	2245.4062	2245.1512	0.2550	0	42	1.1	1	TLVLAMSDPTDLNTVDQVVAK + Oxidation (M)
☑	61	516.3400	1030.6654	1030.5546	0.1108	0	38	5.5	1	VLDELDSLK
☑	76	736.1480	1470.2814	1469.6126	0.6688	0	31	29	1	MGESGDSGAMLLDR + 2 Oxidation (M)
☑	69	692.3760	1382.7374	1382.6388	0.0987	0	28	48	1	EGDLAAAFSMIDK + Oxidation (M)
☑	232	747.0540	2238.1402	2237.0569	1.0833	1	28	40	1	MRMAAEPGAAVTGFGGVDDVVR + 2 Oxidation (M)
☑	226	738.0380	2211.0922	2210.0968	0.9954	1	28	42	1	KYVTTVAAQDATGNQFANSPK
☑	199	1011.1660	3030.4762	3029.5433	0.9329	1	28	63	1	GLVEKGWQQALPDHAAPYTSTMVFLVR + Oxidation (M)
☑	64	633.3340	1264.6534	1265.5267	-0.8733	0	27	50	1	QEMDMENEIK
☑	229	742.4370	2224.2892	2223.1456	1.1435	1	27	39	1	YLKTMGALIEGEETGEITIR
☑	129	876.6700	2626.9882	2627.3894	-0.4012	2	27	62	1	MYRVVSQFNDGIRTPIAQIYIK + Oxidation (M)
☑	190	987.5410	2959.6012	2960.5650	-0.9638	1	27	86	1	VAARLFGGGVVTAVSVPAVMCVLTVVACR
☑	244	892.6220	2674.8442	2675.2757	-0.4315	2	27	29	1	MYGIKGTAKASDDGIYPMMAQNGIK + Oxidation (M)
☑	188	982.2840	2943.8302	2943.4470	0.3832	1	27	63	1	GGVHFLLEVDTPAALEQRMEVASAEMK + Oxidation (M)
☑	98	805.0860	2412.2362	2411.0984	1.1377	1	27	90	1	DLTLLSDYERHEMIEEQMK + 2 Oxidation (M)
☑	155	914.2040	2739.5902	2740.4653	-0.8751	2	26	91	1	AALTLMQIEDLNKTRLANAAAAAAQR + Oxidation (M)
☑	118	853.0970	2556.2692	2557.2119	-0.9427	0	25	1.2e+02	1	SGLDVDGFSVAVCEEIGVPLSNHR
☑	72	701.8480	1401.6814	1402.6690	-0.9875	0	25	1.5e+02	1	LEELSMYDLYK
☑	235	757.8950	2270.6632	2271.1318	-0.4686	0	25	53	1	VLAAPTAIAGFDNGHVMTETTR
☑	228	1111.6590	3331.9552	3331.5602	0.3950	2	25	1.1e+02	1	SYFPVLDLQGHSDCDDVVHMREGLKEIR + Oxidation (M)
☑	196	1006.5340	2011.0534	2009.9782	1.0753	2	25	1.5e+02	1	GKTGLYYFCGRGQSFLR
☑	187	979.3990	1956.7834	1957.0190	-0.2355	0	24	79	1	ALVLQVLDIIDNEEAMR + Oxidation (M)
☑	130	876.9690	2627.8852	2627.3317	0.5535	2	24	1.6e+02	1	HGDNVTNFLKTANHVSLYARGASR
☑	214	695.0820	2082.2242	2082.9536	-0.7295	0	24	92	1	MHNLYGNGSAGAVMIMLEK + 3 Oxidation (M)
☑	206	1029.2490	2056.4834	2055.9944	0.4891	0	24	1.4e+02	1	MMFGLLWMLAASVQAAER + 2 Oxidation (M)
☑	75	734.9050	2201.6932	2201.1362	0.5570	1	24	1.9e+02	1	DTALAQIEGMLKATGPEVVSR + Oxidation (M)
☑	74	731.2890	2190.8452	2191.1307	-0.2855	1	24	1.9e+02	1	SVVFGQDPAIDALASAIKMSR + Oxidation (M)
☑	126	867.5370	2599.5892	2599.3581	0.2311	2	23	1.8e+02	1	SRIGVCKVDITPPIGIDFVGYHR
☑	200	1012.5140	3034.5202	3033.3538	1.1664	0	23	2.1e+02	1	DLHALMTDSQDWWPADFGHYGGLMVR + Oxidation (M)

http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be aware that the annotation is done automatically with little or no manual curation.
Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter cancerogenus, not the sequenced strain), is a member of the Proteobacteria division of the domain bacteria and has been isolated from human feces. The sequenced strain was obtained from ATCC (ATCC 35316).

We have performed one round of automated sequence improvement (pre-finishing), along with manual improvement that includes breaking apart any mis-assembly, and making manual joins where possible. Manual edits also are made where the consensus appears to be incorrect. All low quality data on the ends of contigs is removed. Contigs are ordered and oriented where possible.

This work was supported by the National Human Genome Research Institute and the NIH Roadmap Human Microbiome Project [http://nihroadmap.nih.gov/hmp/].

This is a reference genome for the Human Microbiome Project.
This project is co-owned with the Human Microbiome Project DACC.
Genome Coverage: 29.7x
Sequencing Technology: 454.
Method: conceptual translation.

```
FEATURES             Location/Qualifiers
     source          1..350
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..350
                     /product="outer membrane protein A"
                     /calculated_mol_wt=37337
     Region          1..346
                     /region_name="PRK10808"
                     /note="outer membrane protein A; Reviewed; PRK10808"
                     /db_xref="CDD:138188"
     Region          23..196
                     /region_name="Surface_Ag_2"
                     /note="Surface antigen; cl01155"
                     /db_xref="CDD:141128"
     Region          224..338
                     /region_name="OmpA_C-like"
                     /note="Peptidoglycan binding domains similar to the
                     C-terminal domain of outer-membrane protein OmpA; cd07185"
                     /db_xref="CDD:143586"
     Site            order(229..230,265..266,269,273..274,277,328,332)
                     /site_type="other"
                     /note="ligand binding site"
                     /db_xref="CDD:143586"
     CDS             1..350
                     /locus_tag="EcanA3_020100002425"
                     /coded_by="complement(NZ_ABWM02000004.1:545400..546452)"
                     /note="COG2885 Outer membrane protein and related
                     peptidoglycan-associated (lipo)proteins"
                     /transl_table=11
                     /db_xref="CDD:138188"
```

Mascot: http://www.matrixscience.com/

Mascot Search Results

Protein View

Match to: **gi|261342318** Score: **126**
hypothetical protein EcanA3_17757 [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\280808\SSP 5003_RJ6_01_513.d\SSP 5003_RJ6_01_513.mgf

Nominal mass (M_r): **26485**; Calculated pI value: **5.95**
NCBI BLAST search of gi|261342318 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **13%**

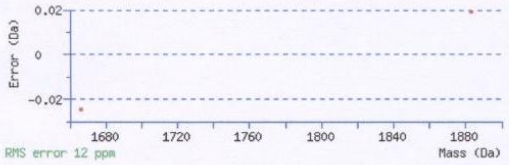
Matched peptides shown in **Bold Red**

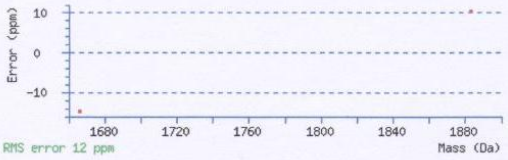
```
  1 MKFNVIALAA LVSLGAVSAQ ANELPAGPHI VTSGTASVDA VPDVATLAIE
 51 VNVAAKDAAS AKKQADDRVA QYLSFLEQNG VAKKDINSAN LRTQPDYDYQ
101 NGKSILKGYR AVRTVEVTLR QLDKLNSLLD GALKAGLNEI RSVSLGVAQP
151 EKYKDEARKA AIDDAIHQAQ QLASGFKGKL GPVYSVRYHV SNYQPSPMVR
201 MMKADAAPVS AQETYEQPTI QFDDQVDVVF QLEPTQTQQT EAAKAQ
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
69 - 83	556.2900	1665.8482	1665.8726	-0.0244	0	R.VAQYLSFLEQNGVAK.K (Ions score 36)
160 - 177	628.6650	1882.9732	1882.9537	0.0195	0	K.AAIDDAIHQAQQLASGFK.G (Ions score 90)





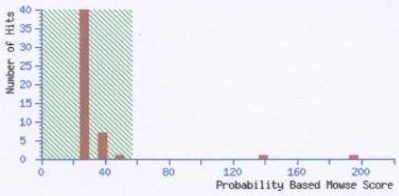

```
LOCUS       ZP_05970176              246 aa            linear   BCT 15-OCT-2009
DEFINITION  hypothetical protein EcanA3_17757 [Enterobacter cancerogenus ATCC
            35316].
ACCESSION   ZP_05970176
VERSION     ZP_05970176.1  GI:261342318
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000033.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 246)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 246)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000033.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
```

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 5102
MS data file    : D:\Data\Lakshmy\120808\260808\SSP 5102_RG17_01_398.d\SSP 5102_RG17_01_398.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:03:53 GMT
Protein hits    : gi|152972368 universal stress protein A [Klebsiella pneumoniae subsp. pneumoniae MGH 78578]
                  gi|50119012  universal stress protein A [Pectobacterium atrosepticum SCRI1043]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|152972368 **Mass:** 16229 **Score:** 195 **Queries matched:** 7 **emPAI:** 0.46
universal stress protein A [Klebsiella pneumoniae subsp. pneumoniae MGH 78578]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	96	711.3850	1420.7554	1420.7926	-0.0371	0	(93)	1.6e-05	1	K.HILIAVDLSPESK.V
☑	97	711.4060	1420.7974	1420.7926	0.0049	0	(82)	0.00018	1	K.HILIAVDLSPESK.V
☑	98	711.4080	1420.8014	1420.7926	0.0089	0	94	1.2e-05	1	K.HILIAVDLSPESK.V
	212	718.3550	2152.0432	2152.8983	-0.8551	0	25	86	2	K.YDMDLVVCGHHQDFWSK.L + Oxidation (M)
☑	111	761.3180	2280.9322	2280.9932	-0.0611	1	20	4.2e+02	1	K.KYDMDLVVCGHHQDFWSK.L + Oxidation (M)
☑	237	826.7860	2477.3362	2477.2472	0.0890	1	56	0.051	1	R.QLINTVHVDMLIVPLRDEEDE.-
☑	240	832.0960	2493.2662	2493.2421	0.0241	1	(51)	0.19	1	R.QLINTVHVDMLIVPLRDEEDE.- + Oxidation (M)

Proteins matching the same set of peptides:
gi|261341988 **Mass:** 16201 **Score:** 195 **Queries matched:** 7
universal stress protein A [Enterobacter cancerogenus ATCC 35316]

---

2. gi|50119012 **Mass:** 16177 **Score:** 134 **Queries matched:** 5 **emPAI:** 0.21
universal stress protein A [Pectobacterium atrosepticum SCRI1043]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
96	711.3850	1420.7554	1420.7926	-0.0371	0	(93)	1.6e-05	1	K.HILIAVDLSPESK.V
97	711.4060	1420.7974	1420.7926	0.0049	0	(82)	0.00018	1	K.HILIAVDLSPESK.V
98	711.4080	1420.8014	1420.7926	0.0089	0	94	1.2e-05	1	K.HILIAVDLSPESK.V
237	826.7860	2477.3362	2476.2632	1.0730	1	(34)	8.2	2	R.QLINTVHVDMLIVPLRDEENE.-
240	832.0960	2493.2662	2492.2581	1.0081	1	40	2.1	2	R.QLINTVHVDMLIVPLRDEENE.- + Oxidation (M)

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	128	836.9540	1671.8934	1671.9858	-0.0924	1	37	5.7	1	MKVVSIVGARPQFIK
☑	230	767.0360	2298.0862	2299.1961	-1.1099	1	36	5.7	1	YLITAAFTDRGGGFNNQTILK
☑	239	831.7950	2492.3632	2492.3097	0.0535	2	35	7.6	1	YPFLLVDRILEMDRENQTIK
☑	249	983.8360	2948.4862	2949.4001	-0.9139	0	34	7.6	1	NSIIYYVGPCPPKPGQVVGSCGPTTSGR
☑	87	651.8450	1301.6754	1301.6914	-0.0160	1	34	13	1	RAFMHTLAEVK
☑	238	826.8270	2477.4592	2476.3471	1.1120	2	33	8.9	1	MLLAEKNVSVLNTAIVERAAYR + Oxidation (M)
☑	233	802.1120	2403.3142	2403.3413	-0.0272	1	32	13	1	KGYVLTGVPNSLSPTQIELFLK
☑	218	737.7020	2210.0842	2211.1787	-1.0945	1	31	17	1	VNVTYEVLDSNVTFKVSVAK
☑	221	742.3240	2223.9502	2223.0841	0.8660	2	30	20	1	SAGVDDIQMEIDEKPINRK + Oxidation (M)
☑	200	702.0780	2103.2122	2103.0406	0.1716	0	29	26	1	ADVVLMVIDATEGVTEQDAK
☑	126	831.3840	2491.1302	2490.2875	0.8426	2	29	60	1	MKQPQGAIFDMLLGKAAVTNWR + Oxidation (M)
☑	214	1091.3100	3270.9082	3271.6078	-0.6996	2	28	44	1	GGHCRLAADVGDEATVRQLYMPDTAVLVTR

```
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..246
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..246
                     /product="hypothetical protein"
                     /calculated_mol_wt=26370
     Region          22..234
                     /region_name="DUF541"
                     /note="Protein of unknown function (DUF541); c101077"
                     /db_xref="CDD:141087"
     CDS             1..246
                     /locus_tag="EcanA3_020100017757"
                     /coded_by="complement(NZ_ABWM02000033.1:268250..268990)"
                     /note="COG2968 Uncharacterized conserved protein"
                     /transl_table=11
                     /db_xref="CDD:138394"
```

Mascot: http://www.matrixscience.com/

Mascot Search Results

Protein View

Match to: **gi|261341988** Score: **195**
universal stress protein A [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\260808\SSP 5102_RG17_01_398.d\SSP 5102_RG17_01_398.mgf

Nominal mass (M_r): **16201**; Calculated pI value: **4.98**
NCBI BLAST search of gi|261341988 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **35%**

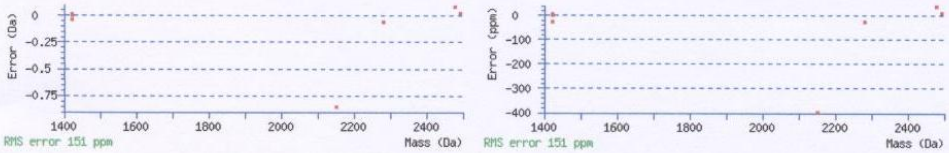
Matched peptides shown in **Bold Red**

```
  1 MAYKHILIAV DLSPESKVLV DKAVSMARPY NAKVSLIHVD VNYSDLYTGL
 51 IDVNLGDMQK RISEETHHAL SELSTNAGYP ITETLSGSGD LGQVLVDAIK
101 KYDMDLVVCG HHQDFWSKLM SSARQLINTV HVDMLIVPLR DEEDE
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
5 - 17	711.3850	1420.7554	1420.7926	-0.0371	0	K.HILIAVDLSPESK.V (Ions score 93)
5 - 17	711.4060	1420.7974	1420.7926	0.0049	0	K.HILIAVDLSPESK.V (Ions score 82)
5 - 17	711.4080	1420.8014	1420.7926	0.0089	0	K.HILIAVDLSPESK.V (Ions score 94)
101 - 118	761.3180	2280.9322	2280.9932	-0.0611	1	K.KYDMDLVVCGHHQDFWSK.L Oxidation (M) (Ions score 20)
102 - 118	718.3550	2152.0432	2152.8983	-0.8551	0	K.YDMDLVVCGHHQDFWSK.L Oxidation (M) (Ions score 25)
125 - 145	826.7860	2477.3362	2477.2472	0.0890	1	R.QLINTVHVDMLIVPLRDEEDE.- (Ions score 56)
125 - 145	832.0960	2493.2662	2493.2421	0.0241	1	R.QLINTVHVDMLIVPLRDEEDE.- Oxidation (M) (Ions score 51)




```
LOCUS       ZP_05969846              145 aa            linear   BCT 15-OCT-2009
DEFINITION  universal stress protein A [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05969846
VERSION     ZP_05969846.1  GI:261341988
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000029.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 145)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 145)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000029.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
```

```
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..145
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..145
                     /product="universal stress protein A"
                     /calculated_mol_wt=16022
     Region          5..137
                     /region_name="USP_Like"
                     /note="Usp: Universal stress protein family. The universal
                     stress protein Usp is a small cytoplasmic bacterial
                     protein whose expression is enhanced when the cell is
                     exposed to stress agents. Usp enhances the rate of cell
                     survival during prolonged exposure to...; cd00293"
                     /db_xref="CDD:30165"
     Site            order(9..11,39,109..110,111..112,121..123)
                     /site_type="other"
                     /note="Ligand Binding Site"
                     /db_xref="CDD:30165"
     CDS             1..145
                     /locus_tag="EcanA3_020100016092"
                     /coded_by="NZ_ABWM02000029.1:151914..152351"
                     /note="COG0589 Universal stress protein UspA and related
                     nucleotide-binding proteins"
                     /transl_table=11
                     /db_xref="CDD:137695"
```

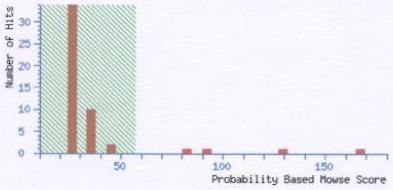
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User          : Lakshmy Manickan
Email         : lakshmy.manickan@unn.ac.uk
Search title  : SSP 5306
MS data file  : D:\Data\Lakshmy\120808\260808\SSP 5306_RG13_01_389.d\SSP 5306_RG13_01_389.mgf
Database      : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy      : Bacteria (Eubacteria) (5690016 sequences)
Timestamp     : 5 Jan 2010 at 11:29:43 GMT
Protein hits  : gi|157155787 periplasmic protein [Escherichia coli E24377A]
                gi|146310199 periplasmic protein [Enterobacter sp. 638]
                gi|238794784 Osmotically-inducible protein Y [Yersinia intermedia ATCC 29909]
                gi|16763356  periplasmic protein [Salmonella enterica subsp. enterica serovar Typhi str. CT18]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ Error tolerant

1. gi|157155787 **Mass:** 21075 **Score:** 167 **Queries matched:** 6 **emPAI:** 0.56
periplasmic protein [Escherichia coli E24377A]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	78	549.7840	1097.5534	1097.6081	-0.0546	0	(62)	0.016	1	K.LLADDIVPSR.H
☑	79	549.8330	1097.6514	1097.6081	0.0434	0	65	0.0095	1	K.LLADDIVPSR.H
☑	80	550.3010	1098.5874	1097.6081	0.9794	0	(46)	1.5	1	K.LLADDIVPSR.H
	96	692.8440	1383.6734	1383.6340	0.0394	0	20	3.2e+02	3	K.VGNFMDDSAITAK.V + Oxidation (M)
☑	188	1011.5000	2020.9854	2021.0317	-0.0463	0	83	0.00013	1	K.VVTLSGFVESQTQAEEAVK.V
☑	189	674.6710	2020.9912	2021.0317	-0.0405	0	(80)	0.00025	1	K.VVTLSGFVESQTQAEEAVK.V

2. gi|146310199 **Mass:** 21390 **Score:** 130 **Queries matched:** 4 **emPAI:** 0.34
periplasmic protein [Enterobacter sp. 638]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	78	549.7840	1097.5534	1097.6081	-0.0546	0	(62)	0.016	1	K.LLADDIVPSR.K
	79	549.8330	1097.6514	1097.6081	0.0434	0	65	0.0095	1	K.LLADDIVPSR.K
	80	550.3010	1098.5874	1097.6081	0.9794	0	(46)	1.5	1	K.LLADDIVPSR.K
☑	96	692.8440	1383.6734	1383.6340	0.0394	0	65	0.0098	1	K.VGNFMDDSSITAK.V

Proteins matching the same set of peptides:
gi|261339031 **Mass:** 21366 **Score:** 130 **Queries matched:** 4
periplasmic protein [Enterobacter cancerogenus ATCC 35316]

3. gi|238794784 **Mass:** 21354 **Score:** 94 **Queries matched:** 4 **emPAI:** 0.16
Osmotically-inducible protein Y [Yersinia intermedia ATCC 29909]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	78	549.7840	1097.5534	1097.6081	-0.0546	0	(62)	0.016	1	K.LLADDIVPSR.K
	79	549.8330	1097.6514	1097.6081	0.0434	0	65	0.0095	1	K.LLADDIVPSR.K
	80	550.3010	1098.5874	1097.6081	0.9794	0	(46)	1.5	1	K.LLADDIVPSR.K
☑	118	781.5810	2341.7212	2341.1510	0.5702	1	29	45	1	K.VETTDGVVQLSGHVESKAQSDR.A

4. gi|16763356 **Score:** 84 **Queries matched:** 4
periplasmic protein [Salmonella enterica subsp. enterica serovar Typhi str. CT18]
☐ Check to include this hit in error tolerant search

Mascot Search Results

Protein View

Match to: **gi|261339031** Score: **130**
periplasmic protein [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\260808\SSP 5306_RG13_01_389.d\SSP 5306_RG13_01_389.mgf

Nominal mass (M_r): **21366**; Calculated pI value: **7.79**
NCBI BLAST search of gi|261339031 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **11%**

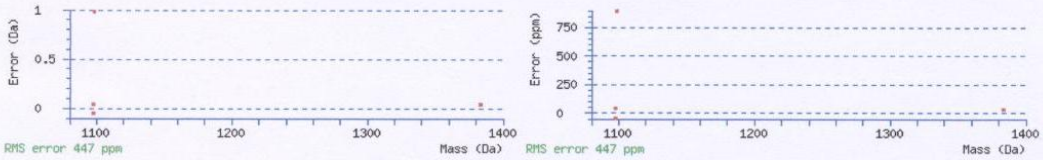
Matched peptides shown in **Bold Red**

```
  1 MNMTRLKISK TLLAVTLGSV LVSGSALAEN TTMDKAQSTA DTAGDKIDSS
 51 MNKVGNFMDD SSITAKVKAA LVDDEAIKST DISVKTDKKV VTLSGFVESQ
101 AQAEQAVKVA KGVEGVTSVS DKLHVRDSKN ASVKGYAGDA ATTSEIKAKL
151 LADDIVPSRK VKVETTDGVV QLSGTVDSQA QSDRAESIAK AIDGVKSVKN
201 DLKTK
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence	
54 - 66	692.8440	1383.6734	1383.6340	0.0394	0	K.VGNFMDDSSITAK.V	(Ions score 65)
150 - 159	549.7840	1097.5534	1097.6081	-0.0546	0	K.LLADDIVPSR.K	(Ions score 62)
150 - 159	549.8330	1097.6514	1097.6081	0.0434	0	K.LLADDIVPSR.K	(Ions score 65)
150 - 159	550.3010	1098.5874	1097.6081	0.9794	0	K.LLADDIVPSR.K	(Ions score 46)




```
LOCUS       ZP_05966889              205 aa            linear   BCT 15-OCT-2009
DEFINITION  periplasmic protein [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05966889
VERSION     ZP_05966889.1  GI:261339031
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000004.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 205)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 205)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000004.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
```

```
          aware that the annotation is done automatically with little or no
          manual curation.
          Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
          gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
          cancerogenus, not the sequenced strain), is a member of the
          Proteobacteria division of the domain bacteria and has been
          isolated from human feces. The sequenced strain was obtained from
          ATCC (ATCC 35316).

          We have performed one round of automated sequence improvement
          (pre-finishing), along with manual improvement that includes
          breaking apart any mis-assembly, and making manual joins where
          possible. Manual edits also are made where the consensus appears to
          be incorrect. All low quality data on the ends of contigs is
          removed. Contigs are ordered and oriented where possible.

          This work was supported by the National Human Genome Research
          Institute and the NIH Roadmap Human Microbiome Project
          [http://nihroadmap.nih.gov/hmp/].

          This is a reference genome for the Human Microbiome Project.
          This project is co-owned with the Human Microbiome Project DACC.
          Genome Coverage: 29.7x
          Sequencing Technology: 454.
          Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..205
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..205
                     /product="periplasmic protein"
                     /calculated_mol_wt=21248
     Region          3..205
                     /region_name="PRK10568"
                     /note="periplasmic protein; Provisional; PRK10568"
                     /db_xref="CDD:137994"
     Region          64..127
                     /region_name="BON"
                     /note="Putative phospholipid-binding domain; cl02771"
                     /db_xref="CDD:141694"
     Region          143..203
                     /region_name="BON"
                     /note="Putative phospholipid-binding domain; cl02771"
                     /db_xref="CDD:141694"
     CDS             1..205
                     /locus_tag="EcanA3_020100001160"
                     /coded_by="complement(NZ_ABWM02000004.1:262747..263364)"
                     /note="COG2823 Predicted periplasmic or secreted
                     lipoprotein"
                     /transl_table=11
                     /db_xref="CDD:137994"
```

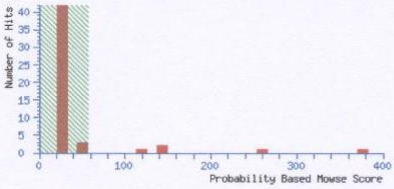
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 5601
MS data file    : D:\Data\Lakshmy\120808\260808\SSP 5601_RG24_01_409.d\SSP 5601_RG24_01_409.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 5 Jan 2010 at 11:30:58 GMT
Protein hits    : gi|261338832 elongation factor Ts [Enterobacter cancerogenus ATCC 35316]
                  gi|1942722   Chain B, Elongation Factor Complex Ef-TuEF-Ts From Escherichia Coli
                  gi|259907543 Elongation factor Ts [Erwinia pyrifoliae Ep1/96]
                  gi|50119972  elongation factor Ts [Pectobacterium atrosepticum SCRI1043]
                  gi|156935315 elongation factor Ts [Enterobacter sakazakii ATCC BAA-894]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|261338832 **Mass:** 30472 **Score:** 376 **Queries matched:** 8 **emPAI:** 0.86
elongation factor Ts [Enterobacter cancerogenus ATCC 35316]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	91	742.8650	1483.7154	1483.6864	0.0290	0	108	4.8e-07	1	K.VETDFAAEVAAMSK.Q + Oxidation (M)
☑	133	887.9130	1773.8114	1773.8646	-0.0531	0	87	4.9e-05	1	R.VSSLEGDVLGSYQHGAR.I
☑	134	592.2800	1773.8182	1773.8646	-0.0464	0	(76)	0.00064	1	R.VSSLEGDVLGSYQHGAR.I
	151	617.6220	1849.8442	1849.9244	-0.0802	0	32	14	5	K.EYQVQLDIAMQSGKPK.E + Oxidation (M)
☑	155	938.4440	1874.8734	1874.9043	-0.0309	0	91	2e-05	1	K.ALTEANGDIELAIENMR.K + Oxidation (M)
☑	156	626.3090	1875.9052	1874.9043	1.0008	0	(33)	14	1	K.ALTEANGDIELAIENMR.K + Oxidation (M)
☑	167	652.6170	1954.8292	1954.9347	-0.1055	0	(41)	1.9	1	K.FTGEVSLTGQPFVMDPSK.S + Oxidation (M)
☑	168	978.4730	1954.9314	1954.9347	-0.0032	0	57	0.048	1	K.FTGEVSLTGQPFVMDPSK.S + Oxidation (M)

2. gi|1942722 **Mass:** 30389 **Score:** 267 **Queries matched:** 6 **emPAI:** 0.23
Chain B, Elongation Factor Complex Ef-TuEF-Ts From Escherichia Coli
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	91	742.8650	1483.7154	1483.6864	0.0290	0	108	4.8e-07	1	K.VETDFAAEVAAMSK.Q + Oxidation (M)
	151	617.6220	1849.8442	1849.9244	-0.0802	0	32	14	5	K.EYQVQLDIAMQSGKPK.E + Oxidation (M)
	155	938.4440	1874.8734	1874.9043	-0.0309	0	91	2e-05	1	K.ALTEANGDIELAIENMR.K + Oxidation (M)
	156	626.3090	1875.9052	1874.9043	1.0008	0	(33)	14	1	K.ALTEANGDIELAIENMR.K + Oxidation (M)
☑	221	757.0520	2268.1342	2269.0936	-0.9595	0	36	6	1	K.IDGNYGIILEVNCQTDFVAK.D
	224	757.7650	2270.2732	2269.0936	1.1795	0	(21)	1.9e+02	3	K.IDGNYGIILEVNCQTDFVAK.D

Proteins matching the same set of peptides:
gi|15799852 **Mass:** 30520 **Score:** 267 **Queries matched:** 6
elongation factor Ts [Escherichia coli O157:H7 EDL933]
gi|157147400 **Mass:** 32239 **Score:** 267 **Queries matched:** 6
elongation factor Ts [Citrobacter koseri ATCC BAA-895]
gi|170768398 **Mass:** 30504 **Score:** 267 **Queries matched:** 6
translation elongation factor Ts [Escherichia albertii TW07627]
gi|172045630 **Mass:** 30471 **Score:** 267 **Queries matched:** 6
RecName: Full=Elongation factor Ts; Short=EF-Ts
gi|194433429 **Mass:** 30538 **Score:** 267 **Queries matched:** 6
translation elongation factor Ts [Shigella dysenteriae 1012]

Mascot Search Results

Protein View

Match to: **gi|261338832** Score: **376**
elongation factor Ts [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\260808\SSP 5601_RG24_01_409.d\SSP 5601_RG24_01_409.mgf

Nominal mass (M_r): **30472**; Calculated pI value: **5.13**
NCBI BLAST search of gi|261338832 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **28%**

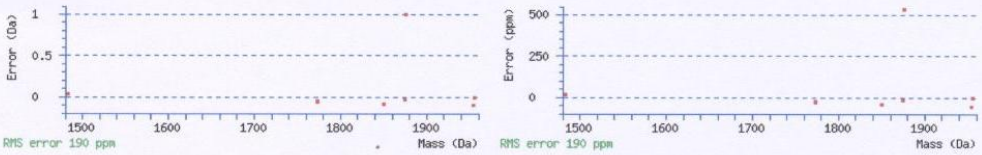
Matched peptides shown in **Bold Red**

```
  1 MAEITASLVK ELRERTGAGM MDCKKALTEA NGDIELAIEN MRKSGAIKAA
 51 KKAGNVAADG VIITKIDGTY GIILEVNCQT DFVAKDGGFQ AFANKVLDAA
101 VAGKITDVEV LKAQFEEERV ALVAKIGENI NIRRVSSLEG DVLGSYQHGA
151 RIGVLVAAKG ADEELVKQLA MHIAASKPEF VKPEDVSAEV VEKEYQVQLD
201 IAMQSGKPKE IAEKMVEGRM KKFTGEVSLT GQPFVMDPSK SVAQLLKEHN
251 ADVTGFIRFE VGEGIEKVET DFAAEVAAMS KQS
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
26 - 42	938.4440	1874.8734	1874.9043	-0.0309	0	K.ALTEANGDIELAIENMR.K Oxidation (M) (Ions score 91)
26 - 42	626.3090	1875.9052	1874.9043	1.0008	0	K.ALTEANGDIELAIENMR.K Oxidation (M) (Ions score 33)
135 - 151	887.9130	1773.8114	1773.8646	-0.0531	0	R.VSSLEGDVLGSYQHGAR.I (Ions score 87)
135 - 151	592.2800	1773.8182	1773.8646	-0.0464	0	R.VSSLEGDVLGSYQHGAR.I (Ions score 76)
194 - 209	617.6220	1849.8442	1849.9244	-0.0802	0	K.EYQVQLDIAMQSGKPK.E Oxidation (M) (Ions score 32)
223 - 240	652.6170	1954.8292	1954.9347	-0.1055	0	K.FTGEVSLTGQPFVMDPSK.S Oxidation (M) (Ions score 41)
223 - 240	978.4730	1954.9314	1954.9347	-0.0032	0	K.FTGEVSLTGQPFVMDPSK.S Oxidation (M) (Ions score 57)
268 - 281	742.8650	1483.7154	1483.6864	0.0290	0	K.VETDFAAEVAAMSK.Q Oxidation (M) (Ions score 108)




```
LOCUS       ZP_05966690              283 aa            linear   BCT 15-OCT-2009
DEFINITION  elongation factor Ts [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05966690
VERSION     ZP_05966690.1  GI:261338832
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000004.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 283)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 283)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000004.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
```

```
          aware that the annotation is done automatically with little or no
          manual curation.
          Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
          gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
          cancerogenus, not the sequenced strain), is a member of the
          Proteobacteria division of the domain bacteria and has been
          isolated from human feces. The sequenced strain was obtained from
          ATCC (ATCC 35316).

          We have performed one round of automated sequence improvement
          (pre-finishing), along with manual improvement that includes
          breaking apart any mis-assembly, and making manual joins where
          possible. Manual edits also are made where the consensus appears to
          be incorrect. All low quality data on the ends of contigs is
          removed. Contigs are ordered and oriented where possible.

          This work was supported by the National Human Genome Research
          Institute and the NIH Roadmap Human Microbiome Project
          [http://nihroadmap.nih.gov/hmp/].

          This is a reference genome for the Human Microbiome Project.
          This project is co-owned with the Human Microbiome Project DACC.
          Genome Coverage: 29.7x
          Sequencing Technology: 454.
          Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..283
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..283
                     /product="elongation factor Ts"
                     /calculated_mol_wt=30244
     Region          1..279
                     /region_name="tsf"
                     /note="elongation factor Ts; Provisional; PRK09377"
                     /db_xref="CDD:137239"
     Region          4..39
                     /region_name="UBA"
                     /note="Ubiquitin Associated domain. The UBA domain is a
                     commonly occurring sequence motif in some members of the
                     ubiquitination pathway, UV excision repair proteins, and
                     certain protein kinases. Although its specific role is so
                     far unknown, it has been...; cl00153"
                     /db_xref="CDD:140477"
     Region          61..263
                     /region_name="EF_TS"
                     /note="Elongation factor TS; pfam00889"
                     /db_xref="CDD:109927"
     CDS             1..283
                     /gene="tsf"
                     /locus_tag="EcanA3_020100000165"
                     /coded_by="complement(NZ_ABWM02000004.1:30889..31740)"
                     /note="COG0264 Translation elongation factor Ts"
                     /transl_table=11
                     /db_xref="CDD:137239"
```

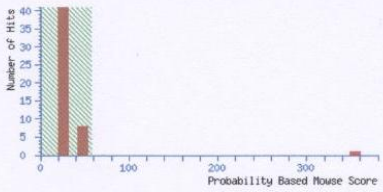
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 5607
MS data file    : D:\Data\Lakshmy\120808\280808\SSP 5607_RH20_01_448.d\SSP 5607_RH20_01_448.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 5 Jan 2010 at 11:32:24 GMT
Protein hits    : gi|261339130 carbamate kinase [Enterobacter cancerogenus ATCC 35316]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|261339130 **Mass:** 32346 **Score:** 355 **Queries matched:** 6 **emPAI:** 0.48
carbamate kinase [Enterobacter cancerogenus ATCC 35316]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	33	465.2430	928.4714	928.5341	-0.0627	0	50	0.4	1	K.DLSAALLAR.Q
☑	63	714.8890	1427.7634	1427.7984	-0.0349	0	86	6.9e-05	1	R.IVESDAITALIQR.D
	65	498.6310	1492.8712	1492.9341	-0.0629	0	27	55	5	R.KPTLVVALGGNALLK.R
☑	98	878.0740	1754.1334	1753.9938	0.1397	1	80	0.0002	1	R.GIEAVIDKDLSAALLAR.Q
☑	181	766.7240	2297.1502	2297.0859	0.0643	0	55	0.075	1	R.DHLVICNGGGGVPVVENANGYR.G
☑	188	798.1100	2391.3082	2391.2910	0.0171	0	57	0.047	1	R.VVLVHGNGPQVGLLALQNSAYDK.V

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	39	476.8730	951.7314	952.5818	-0.8503	1	43	0.96	1	AAPGGAILKR
☑	51	587.8290	1173.6434	1173.7332	-0.0898	2	36	7.2	1	TLKLTKETIK
☑	65	498.6310	1492.8712	1491.7351	1.1360	1	33	12	1	RGAASATGLGEDLMK + Oxidation (M)
☑	104	608.9710	1823.8912	1823.9377	-0.0465	1	33	12	1	QELADEAEIVLPRGER
☑	88	563.1710	1686.4912	1685.9147	0.5764	2	30	18	1	ADRRVLLWLANGMR + Oxidation (M)
☑	49	577.7330	1730.1772	1730.8951	-0.7179	1	30	43	1	LPNIDYRVALAEDSR
☑	55	618.9030	1235.7914	1235.6067	0.1847	1	30	27	1	VEEAMDTVKAK + Oxidation (M)
☑	140	1034.4450	3100.3132	3099.5196	0.7936	1	29	49	1	GHAEAELNIGFTAVASSGPLPDAMAAFRQR + Oxidation (M)
☑	199	873.4630	2617.3672	2617.2417	0.1255	2	28	31	1	MRMTFRWYGENNDSVTLEQIK
☑	122	976.7320	2927.1742	2926.4827	0.6915	2	28	52	1	NPDAAELVRGHSGRIVGCLTALMITMK + Oxidation (M)
☑	90	565.5930	1693.7572	1692.8366	0.9206	2	27	47	1	SGMRSAEPRSLPSFR + Oxidation (M)
☑	109	927.1790	2778.5152	2779.4691	-0.9539	2	27	71	1	QFGKKTAVSSVTFDIPQGQMVGIIGR + Oxidation (M)
☑	139	1034.1020	3099.2842	3098.6652	0.6190	2	27	73	1	ITKLRLYVSGSNLLTFSSLYQGLDPEGK
☑	182	771.7530	2312.2372	2312.1107	0.1265	1	26	54	1	CRSFDAGADGTAFAEGIGVVALK
☑	166	742.3730	2224.0972	2224.0549	0.0423	0	25	70	1	FGPPGASSNGFNENLALFNSGK
☑	124	985.8260	2954.4562	2954.4194	0.0368	0	25	1.1e+02	1	ETGAWASMVFVPPAFVADAVMEAASSGIK + Oxidation (M)
☑	113	944.0300	2829.0682	2829.4669	-0.3988	2	25	1.3e+02	1	GIRHLKLIDLCTIPVSYAWAEAMR + Oxidation (M)
☑	115	947.1350	2838.3832	2838.4182	-0.0350	1	25	1.3e+02	1	TTSDAALSGGPDPAPAPRAILMSGLQGSGK + Oxidation (M)
☑	46	563.4510	1687.3312	1687.8199	-0.4887	0	24	1.3e+02	1	SSMAQAGKPIEPATER + Oxidation (M)
☑	202	883.8770	2648.6092	2648.3129	0.2963	2	24	57	1	LQAVRFMWEGLVAGETGSPGRSER + Oxidation (M)
☑	99	881.5150	2641.5232	2640.3619	1.1612	0	24	1.6e+02	1	FLSQAQLSQAEQINTANAQPNIVR
☑	149	1059.9540	3176.8402	3176.6713	0.1689	1	24	1.3e+02	1	DLLKEIMESAIPVTMQDVVLVFANVSGVK + 2 Oxidation (M)
☑	151	1067.5050	3199.4932	3198.5734	0.9198	2	24	1.5e+02	1	KTEETLSSSYRVQFNFRPVASTADYFR
☑	94	573.3000	1716.8782	1716.9860	-0.1078	2	24	1.2e+02	1	GRLHLDAGAVAAVVGRR
☑	101	599.0420	1794.1042	1794.8788	-0.7746	2	24	99	1	DWIELKADKNASYDK
☑	54	615.6450	1843.9132	1843.8332	0.0800	0	23	2.2e+02	1	VDVYGTMDEASAAMGLAK + Oxidation (M)
☑	152	1067.5820	2133.1494	2133.0815	0.0680	0	23	1.9e+02	1	ANPVTQQPPGSSLGTPSPAATR
☑	117	952.4810	1902.9474	1904.0441	-1.0967	1	23	1.2e+02	1	MTPEQLQGKVLTGVLYK
☑	89	847.3550	2539.0432	2538.4866	0.5566	2	23	2e+02	1	MALVNKVLIIGGGIGGMSAAILLRR + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|261339130** Score: **355**
carbamate kinase [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\280808\SSP 5607_RH20_01_448.d\SSP 5607_RH20_01_448.mgf

Nominal mass (M_r): **32346**; Calculated pI value: **5.13**
NCBI BLAST search of gi|261339130 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **29%**

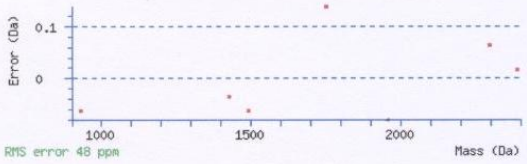
Matched peptides shown in **Bold Red**

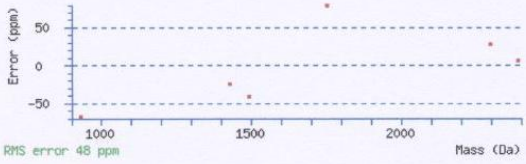
```
  1 MERKPTLVVA LGGNALLKRG EPLEAEIQRQ NIELAARTIA GLTAQWRVVL
 51 VHGNGPQVGL LALQNSAYDK VTPYPLDILG AESQGMIGYM LQQALKNNLP
101 QREVSVLLTQ VEVDPADPAF TNPTKYIGPI YTDAQAKALA AEKGWIFKAD
151 GSAFRRVVPS PQPKRIVESD AITALIQRDH LVICNGGGGV PVVENANGYR
201 GIEAVIDKDL SAALLARQIE ADALLILTDA DAVYLDWGKP TQRPLAQVTP
251 ALLNDMQFDA GSMGPKVAAC REFVEACGGI AGIGALADGA EILAGEKGTL
301 IRN
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
4 - 18	498.6310	1492.8712	1492.9341	-0.0629	0	R.KPTLVVALGGNALLK.R (Ions score 27)
48 - 70	798.1100	2391.3082	2391.2910	0.0171	0	R.VVLVHGNGPQVGLLALQNSAYDK.V (Ions score 57)
166 - 178	714.8890	1427.7634	1427.7984	-0.0349	0	R.IVESDAITALIQR.D (Ions score 86)
179 - 200	766.7240	2297.1502	2297.0859	0.0643	0	R.DHLVICNGGGGVPVVENANGYR.G (Ions score 55)
201 - 217	878.0740	1754.1334	1753.9938	0.1397	1	R.GIEAVIDKDLSAALLAR.Q (Ions score 80)
209 - 217	465.2430	928.4714	928.5341	-0.0627	0	K.DLSAALLAR.Q (Ions score 50)






```
LOCUS       ZP_05966988              303 aa            linear   BCT 15-OCT-2009
DEFINITION  carbamate kinase [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05966988
VERSION     ZP_05966988.1  GI:261339130
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000004.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 303)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 303)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000004.
```

```
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..303
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..303
                     /product="carbamate kinase"
                     /EC_number="2.7.2.2"
                     /calculated_mol_wt=32061
     Region          6..274
                     /region_name="AAK_CK"
                     /note="AAK_CK: Carbamate kinase (CK) catalyzes both the
                     ATP-phosphorylation of carbamate and carbamoyl phosphate
                     (CP) utilization with the production of ATP from ADP and
                     CP. Both CK (this CD) and nonhomologous CP synthetase
                     synthesize carbamoyl phosphate, an...; cd04235"
                     /db_xref="CDD:58601"
     Site            order(10,12..13,53..55,125,207..209)
                     /site_type="other"
                     /note="putative substrate binding site"
                     /db_xref="CDD:58601"
     Site            order(13,228..229,234,237,258,262..263,266)
                     /site_type="other"
                     /note="nucleotide binding site"
                     /db_xref="CDD:58601"
     Site            order(13,228..229,234,237,258,262..263,266)
                     /site_type="other"
                     /note="nucleotide binding site"
                     /db_xref="CDD:58601"
     Site            order(63,74,77..78,81,85..86,89,92..93,97,106..108,110,
                     168,171,199,201)
                     /site_type="other"
                     /note="homodimer interface"
                     /db_xref="CDD:58601"
     CDS             1..303
                     /locus_tag="EcanA3_020100001655"
                     /coded_by="NZ_ABWM02000004.1:365507..366418"
                     /note="COG0549 Carbamate kinase"
                     /transl_table=11
                     /db_xref="CDD:138900"
```

Mascot: http://www.matrixscience.com/

☑	81	754.9170	1507.8194	1507.7518	0.0676	1	67	0.006	1	K.SLYEADLVDEAKR.L
	111	580.6830	1739.0272	1738.9763	0.0508	1	24	1.1e+02	3	K.ALKEPARPMVAIVGGSK.V + Oxidation (M)
	135	938.0120	1874.0094	1873.9462	0.0633	0	97	5.8e-06	1	K.ISYISTGGGAFLEFVEGK.V
	147	665.0230	1992.0472	1992.0258	0.0214	0	55	0.084	1	K.TILWNGPVGVFEFPNFR.K
	162	687.3070	2058.8992	2059.9132	-1.0140	0	49	0.31	1	K.YAALCDVFVMDAFGTAHR.A + Oxidation (M)
☑	163	1030.5570	2059.0994	2060.0136	-0.9141	0	(32)	16	1	K.FADVACAGPLLADELEALGK.A
☑	164	687.3740	2059.1002	2060.0136	-0.9134	0	50	0.22	1	K.FADVACAGPLLADELEALGK.A
	233	817.4700	2449.3882	2449.3078	0.0804	0	80	0.00022	1	K.IADQLIVGGGIANTFVAAQGHNVGK.S

3. gi|156932627 **Mass:** 41426 **Score:** 613 **Queries matched:** 11 **emPAI:** 1.00
phosphoglycerate kinase [Enterobacter sakazakii ATCC BAA-894]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
56	578.3620	1154.7094	1154.6910	0.0184	0	47	0.62	1	R.ASLPTIELALK.Q
58	622.3450	1242.6754	1242.6642	0.0113	0	53	0.17	1	K.VLPAVAMLEER.A + Oxidation (M)
65	676.8330	1351.6514	1351.6507	0.0008	0	58	0.042	1	K.SLYEADLVDEAK.R
81	754.9170	1507.8194	1507.7518	0.0676	1	67	0.006	1	K.SLYEADLVDEAKR.L
111	580.6830	1739.0272	1738.9763	0.0508	1	24	1.1e+02	3	K.ALKEPARPMVAIVGGSK.V + Oxidation (M)
135	938.0120	1874.0094	1873.9462	0.0633	0	97	5.8e-06	1	K.ISYISTGGGAFLEFVEGK.V
147	665.0230	1992.0472	1992.0258	0.0214	0	55	0.084	1	K.TILWNGPVGVFEFPNFR.K
163	1030.5570	2059.0994	2060.0136	-0.9141	0	(19)	3.5e+02	7	K.FADVACAGPLLAEELDALGK.A
164	687.3740	2059.1002	2060.0136	-0.9134	0	34	9.3	2	K.FADVACAGPLLAEELDALGK.A
200	1109.6410	2217.2674	2217.1165	0.1510	0	102	1.5e-06	1	K.DYLDGVEVAEGELVVLENVR.F
233	817.4700	2449.3882	2449.3078	0.0804	0	80	0.00022	1	K.IADQLIVGGGIANTFVAAQGHNVGK.S

Proteins matching the same set of peptides:
gi|260599261 **Mass:** 41470 **Score:** 613 **Queries matched:** 11
Phosphoglycerate kinase [Cronobacter turicensis]

4. gi|242238006 **Mass:** 41399 **Score:** 435 **Queries matched:** 7 **emPAI:** 0.54
Phosphoglycerate kinase [Dickeya dadantii Ech703]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
56	578.3620	1154.7094	1154.6910	0.0184	0	47	0.62	1	R.ASLPTIEIALK.Q
72	734.9040	1467.7934	1467.7457	0.0478	0	112	1.9e-07	1	R.VATEFSETATATLK.S
73	490.2880	1467.8422	1467.7457	0.0965	0	(28)	40	1	R.VATEFSETATATLK.S
135	938.0120	1874.0094	1873.9462	0.0633	0	97	5.8e-06	1	K.ISYISTGGGAFLEFVEGK.K
147	665.0230	1992.0472	1992.0258	0.0214	0	55	0.084	1	K.TILWNGPVGVFEFPNFR.K
162	687.3070	2058.8992	2059.9132	-1.0140	0	49	0.31	1	K.YAALCDVFVMDAFGTAHR.A + Oxidation (M)
233	817.4700	2449.3882	2449.3078	0.0804	0	80	0.00022	1	R.IADQLIVGGGIANTFVAAQGHNVGK.S

5. gi|227114292 **Mass:** 41338 **Score:** 362 **Queries matched:** 6 **emPAI:** 0.54
phosphoglycerate kinase [Pectobacterium carotovorum subsp. brasiliensis PBR1692]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
56	578.3620	1154.7094	1154.6910	0.0184	0	47	0.62	1	R.ASLPTIEIALK.Q
135	938.0120	1874.0094	1873.9462	0.0633	0	97	5.8e-06	1	K.ISYISTGGGAFLEFVEGK.K
147	665.0230	1992.0472	1992.0258	0.0214	0	55	0.084	1	K.TILWNGPVGVFEFPNFR.K
162	687.3070	2058.8992	2059.9132	-1.0140	0	49	0.31	1	K.YAALCDVFVMDAFGTAHR.A + Oxidation (M)
233	817.4700	2449.3882	2449.3078	0.0804	0	80	0.00022	1	K.IADQLIVGGGIANTFVAAQGHNVGK.S
☑ 242	903.9790	2708.9152	2708.3545	0.5607	0	38	2.3	1	R.AIAESDAFSIAGGGDTLAAIDLFGIADK.I

Proteins matching the same set of peptides:
gi|253690062 **Mass:** 41334 **Score:** 362 **Queries matched:** 6
Phosphoglycerate kinase [Pectobacterium carotovorum subsp. carotovorum PC1]
gi|261823130 **Mass:** 41340 **Score:** 362 **Queries matched:** 6
Phosphoglycerate kinase [Pectobacterium wasabiae WPP163]

6. gi|157372179 **Mass:** 41484 **Score:** 319 **Queries matched:** 5 **emPAI:** 0.30
phosphoglycerate kinase [Serratia proteamaculans 568]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
58	622.3450	1242.6754	1242.6642	0.0113	0	28	45	2	K.ALPAVVMLEER.A + Oxidation (M)
135	938.0120	1874.0094	1873.9462	0.0633	0	97	5.8e-06	1	K.ISYISTGGGAFLEFVEGK.A
147	665.0230	1992.0472	1992.0258	0.0214	0	55	0.084	1	K.TILWNGPVGVFEFPNFR.K
162	687.3070	2058.8992	2059.9132	-1.0140	0	38	4	2	K.YAALCDVYVMDAFGTAHR.A
200	1109.6410	2217.2674	2217.1165	0.1510	0	102	1.5e-06	1	K.DYLEGVDVAEGELVVLENVR.F

7. gi|238759311 **Mass:** 39537 **Score:** 306 **Queries matched:** 5 **emPAI:** 0.44
Phosphoglycerate kinase [Yersinia aldovae ATCC 35236]
☐ Check to include this hit in error tolerant search

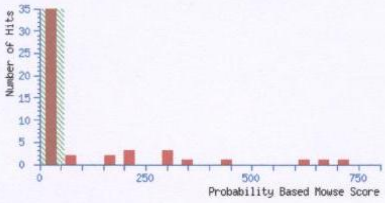
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
135	938.0120	1874.0094	1873.9462	0.0633	0	97	5.8e-06	1	K.ISYISTGGGAFLEFVEGK.K
147	665.0230	1992.0472	1992.0258	0.0214	0	55	0.084	1	K.TILWNGPVGVFEFPNFR.K
162	687.3070	2058.8992	2059.9132	-1.0140	0	38	4	2	K.YAALCDVYVMDAFGTAHR.A
233	817.4700	2449.3882	2449.3078	0.0804	0	80	0.00022	1	K.IADQLIVGGGIANTFVAAQGHNVGK.S
242	903.9790	2708.9152	2708.3545	0.5607	0	38	2.3	1	R.AIAESDAFSIAGGGDTLAAIDLFGIADK.I

Mascot Search Results

```
User           : Lakshmy Manickan
Email          : lakshmy.manickan@unn.ac.uk
Search title   : SSP  5701
MS data file   : D:\Data\Lakshmy\120808\190808\SSP 5701_RF24_01_363.d\SSP 5701_RF24_01_363.mgf
Database       : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy       : Bacteria (Eubacteria) (5690016 sequences)
Timestamp      : 6 Jan 2010 at 13:02:00 GMT
Protein hits   : gi|146312978 phosphoglycerate kinase [Enterobacter sp. 638]
                 gi|261342323 phosphoglycerate kinase [Enterobacter cancerogenus ATCC 35316]
                 gi|156932627 phosphoglycerate kinase [Enterobacter sakazakii ATCC BAA-894]
                 gi|242238006 Phosphoglycerate kinase [Dickeya dadantii Ech703]
                 gi|227114292 phosphoglycerate kinase [Pectobacterium carotovorum subsp. brasiliensis PBR1692]
                 gi|157372179 phosphoglycerate kinase [Serratia proteamaculans 568]
                 gi|238759311 Phosphoglycerate kinase [Yersinia aldovae ATCC 35236]
                 gi|261346331 phosphoglycerate kinase [Providencia rustigianii DSM 4541]
                 gi|268592003 phosphoglycerate kinase [Providencia rettgeri DSM 1131]
                 gi|53733312  COG0126: 3-phosphoglycerate kinase [Haemophilus influenzae R2866]
                 gi|183601149 hypothetical protein PROSTU_04779 [Providencia stuartii ATCC 25827]
                 gi|22127184  phosphoglycerate kinase [Yersinia pestis KIM]
                 gi|220936052 Phosphoglycerate kinase [Thioalkalivibrio sp. HL-EbGR7]
                 gi|15640504  phosphoglycerate kinase [Vibrio cholerae O1 biovar El Tor str. N16961]
                 gi|261210056 phosphoglycerate kinase [Vibrio sp. RC341]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned Decreasing Score Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|146312978 **Mass:** 41414 **Score:** 713 **Queries matched:** 13 **emPAI:** 1.00
phosphoglycerate kinase [Enterobacter sp. 638]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	56	578.3620	1154.7094	1154.6910	0.0184	0	47	0.62	1	R.ASLPTIELALK.Q
☑	58	622.3450	1242.6754	1242.6642	0.0113	0	53	0.17	1	K.VLPAVAMLEER.A + Oxidation (M)
☑	72	734.9040	1467.7934	1467.7457	0.0478	0	112	1.9e-07	1	R.VATEFSETATATLK.S
☑	73	490.2880	1467.8422	1467.7457	0.0965	0	(28)	40	1	R.VATEFSETATATLK.S
	111	580.6830	1739.0272	1738.9763	0.0508	1	24	1.1e+02	3	K.ALKEPARPMVAIVGGSK.V + Oxidation (M)
☑	135	938.0120	1874.0094	1873.9462	0.0633	0	97	5.8e-06	1	K.ISYISTGGGAFLEFVEGK.V
☑	147	665.0230	1992.0472	1992.0258	0.0214	0	55	0.084	1	K.TILWNGPVGVFEFPNFR.K
☑	162	687.3070	2058.8992	2059.9132	-1.0140	0	49	0.31	1	K.YAALCDVFVMDAFGTAHR.A + Oxidation (M)
	163	1030.5570	2059.0994	2060.0136	-0.9141	0	(19)	3.5e+02	7	K.FADVACAGPLLAEELDALGK.A
	164	687.3740	2059.1002	2060.0136	-0.9134	0	34	9.3	2	K.FADVACAGPLLAEELDALGK.A
☑	200	1109.6410	2217.2674	2217.1165	0.1510	0	102	1.5e-06	1	K.DYLDGVEVAEGELVVLENVR.F
☑	220	772.7840	2315.3302	2315.1492	0.1809	1	68	0.0031	1	K.SVNDIKDEEQILDLGDVSAQK.L
☑	233	817.4700	2449.3882	2449.3078	0.0804	0	80	0.00022	1	K.IADQLIVGGGIANTFVAAQGHNVGK.S

2. gi|261342323 **Mass:** 41343 **Score:** 684 **Queries matched:** 13 **emPAI:** 1.19
phosphoglycerate kinase [Enterobacter cancerogenus ATCC 35316]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	56	578.3620	1154.7094	1154.6910	0.0184	0	47	0.62	1	R.ASLPTIELALK.Q
	58	622.3450	1242.6754	1242.6642	0.0113	0	53	0.17	1	K.VLPAVAMLEER.A + Oxidation (M)
☑	65	676.8330	1351.6514	1351.6507	0.0008	0	58	0.042	1	K.SLYEADLVDEAK.R
	72	734.9040	1467.7934	1467.7457	0.0478	0	112	1.9e-07	1	R.VATEFSETATATLK.S
	73	490.2880	1467.8422	1467.7457	0.0965	0	(28)	40	1	R.VATEFSETATATLK.S

Mascot Search Results

Protein View

Match to: **gi|261342323** Score: **684**
phosphoglycerate kinase [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\190808\SSP 5701_RF24_01_363.d\SSP 5701_RF24_01_363.mgf

Nominal mass (M_r): **41343**; Calculated pI value: **5.08**
NCBI BLAST search of gi|261342323 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **42%**

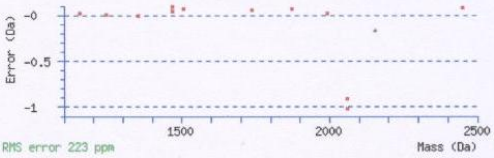
Matched peptides shown in **Bold Red**

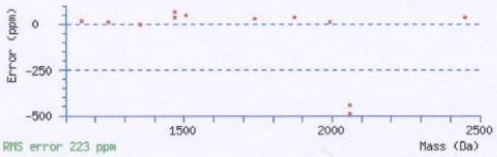
```
  1 MSVIKMTDLD LAGKRVFIRA DLNVPVKDGK VTSDARIRAS LPTIELALKQ
 51 GAKVMVTSHL GRPTEGEYNE EFSLLPVVNY LKDKLSSPVR LVKDYLDGVE
101 VAAGELVVLE NVRFNKGEKK DDETLSKKYA ALCDVFVMDA FGTAHRAQAS
151 THGIGKFADV ACAGPLLADE LEALGKALKE PARPMVAIVG GSKVSTKLTV
201 LDSLSKIADQ LIVGGGIANT FVAAQGHNVG KSLYEADLVD EAKRLLTTCD
251 IPVPTDVRVA TEFSETATAT LKSVNDIKDD EQILDLGDVS AQKLAEILKN
301 AKTILWNGPV GVFEFPNFRK GTEIVANAIA DSEAFSIAGG GDTLAAIDLF
351 GIADKISYIS TGGGAFLEFV EGKVLPAVAM LEERAKQ
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
39 - 49	578.3620	1154.7094	1154.6910	0.0184	0	R.ASLPTIELALK.Q (Ions score 47)
129 - 146	687.3070	2058.8992	2059.9132	-1.0140	0	K.YAALCDVFVMDAFGTAHR.A Oxidation (M) (Ions score 49)
157 - 176	1030.5570	2059.0994	2060.0136	-0.9141	0	K.FADVACAGPLLADELEALGK.A (Ions score 32)
157 - 176	687.3740	2059.1002	2060.0136	-0.9134	0	K.FADVACAGPLLADELEALGK.A (Ions score 50)
177 - 193	580.6830	1739.0272	1738.9763	0.0508	1	K.ALKEPARPMVAIVGGSK.V Oxidation (M) (Ions score 24)
207 - 231	817.4700	2449.3882	2449.3078	0.0804	0	K.IADQLIVGGGIANTFVAAQGHNVGK.S (Ions score 80)
232 - 243	676.8330	1351.6514	1351.6507	0.0008	0	K.SLYEADLVDEAK.R (Ions score 58)
232 - 244	754.9170	1507.8194	1507.7518	0.0676	1	K.SLYEADLVDEAKR.L (Ions score 67)
259 - 272	734.9040	1467.7934	1467.7457	0.0478	0	R.VATEFSETATATLK.S (Ions score 112)
259 - 272	490.2880	1467.8422	1467.7457	0.0965	0	R.VATEFSETATATLK.S (Ions score 28)
303 - 319	665.0230	1992.0472	1992.0258	0.0214	0	K.TILWNGPVGVFEFPNFR.K (Ions score 55)
356 - 373	938.0120	1874.0094	1873.9462	0.0633	0	K.ISYISTGGGAFLEFVEGK.V (Ions score 97)
374 - 384	622.3450	1242.6754	1242.6642	0.0113	0	K.VLPAVAMLEER.A Oxidation (M) (Ions score 53)






```
LOCUS       ZP_05970181              387 aa            linear   BCT 15-OCT-2009
DEFINITION  phosphoglycerate kinase [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05970181
VERSION     ZP_05970181.1  GI:261342323
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000033.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 387)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 387)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
```

```
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000033.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..387
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..387
                     /product="phosphoglycerate kinase"
                     /EC_number="2.7.2.3"
                     /calculated_mol_wt=41063
     Region          6..383
                     /region_name="Phosphoglycerate_kinase"
                     /note="Phosphoglycerate kinase (PGK) is a monomeric enzyme
                     which catalyzes the transfer of the high-energy phosphate
                     group of 1,3-bisphosphoglycerate to ADP, forming ATP and
                     3-phosphoglycerate. This reaction represents the first of
                     the two substrate-level...; cd00318"
                     /db_xref="CDD:29400"
     Site            order(21,23,36,59,113)
                     /site_type="other"
                     /note="substrate binding site"
                     /db_xref="CDD:29400"
     Site            order(178..181,359..361)
                     /site_type="other"
                     /note="hinge regions"
                     /db_xref="CDD:29400"
     Site            order(216,283,307,309,311..314,341..343)
                     /site_type="other"
                     /note="ADP binding site"
                     /db_xref="CDD:29400"
     Site            342
                     /site_type="other"
                     /note="catalytic site"
                     /db_xref="CDD:29400"
     CDS             1..387
                     /gene="pgk"
                     /locus_tag="EcanA3_020100017782"
                     /coded_by="complement(NZ_ABWM02000033.1:273242..274405)"
                     /note="COG0126 3-phosphoglycerate kinase"
                     /transl_table=11
                     /db_xref="CDD:134081"
```

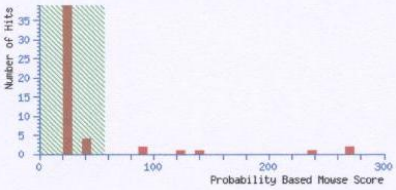
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 5702
MS data file    : D:\Data\Lakshmy\120808\280808\SSP 5702_RI12_01_475.d\SSP 5702_RI12_01_475.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 5 Jan 2010 at 11:34:50 GMT
Protein hits    : gi|146312874 phosphopyruvate hydratase [Enterobacter sp. 638]
                  gi|156932734 phosphopyruvate hydratase [Enterobacter sakazakii ATCC BAA-894]
                  gi|157148317 phosphopyruvate hydratase [Citrobacter koseri ATCC BAA-895]
                  gi|260771907 enolase [Vibrio metschnikovii CIP 69.14]
                  gi|198241971 phosphopyruvate hydratase [Salmonella enterica subsp. enterica serovar Dublin str. CT_02021853]
                  gi|15642443  phosphopyruvate hydratase [Vibrio cholerae O1 biovar El Tor str. N16961]
                  gi|54310177  phosphopyruvate hydratase [Photobacterium profundum SS9]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|146312874 **Mass:** 45497 **Score:** 270 **Queries matched:** 4 **emPAI:** 0.15
phosphopyruvate hydratase [Enterobacter sp. 638]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	68	478.8270	955.6394	955.6066	0.0328	0	45	0.95	1	K.GIVNSILIK.F
	105	733.3860	1464.7574	1463.8460	0.9114	0	55	0.091	2	K.AVGAVNGPIAQAVLGK.D
☑	117	781.9560	1561.8974	1561.8352	0.0623	0	92	2e-05	1	K.IQLVGDDLFVTNTK.I
☑	177	957.9100	1913.8054	1913.9047	-0.0993	0	79	0.00031	1	K.AFTSEEFTHFLEDLTK.Q

2. gi|156932734 **Mass:** 45573 **Score:** 262 **Queries matched:** 4 **emPAI:** 0.23
phosphopyruvate hydratase [Enterobacter sakazakii ATCC BAA-894]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	105	733.3860	1464.7574	1463.8460	0.9114	0	69	0.0032	1	K.AVGAVNGPIAQAIVGK.D
	117	781.9560	1561.8974	1561.8352	0.0623	0	92	2e-05	1	K.IQLVGDDLFVTNTK.I
	148	875.8420	1749.6694	1750.7430	-1.0735	0	22	1.2e+02	5	K.DITLAMDCAASEFYK.D + Oxidation (M)
	177	957.9100	1913.8054	1913.9047	-0.0993	0	79	0.00031	1	K.AFTSEEFTHFLEDLTK.Q

Proteins matching the same set of peptides:
gi|260599132 **Mass:** 45516 **Score:** 262 **Queries matched:** 4
Enolase [Cronobacter turicensis]

3. gi|157148317 **Mass:** 45646 **Score:** 238 **Queries matched:** 4 **emPAI:** 0.15
phosphopyruvate hydratase [Citrobacter koseri ATCC BAA-895]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
68	478.8270	955.6394	955.6066	0.0328	0	45	0.95	1	K.GIVNSILIK.F
117	781.9560	1561.8974	1561.8352	0.0623	0	92	2e-05	1	K.IQLVGDDLFVTNTK.I
148	875.8420	1749.6694	1750.7430	-1.0735	0	22	1.2e+02	5	K.DITLAMDCAASEFYK.D + Oxidation (M)
177	957.9100	1913.8054	1913.9047	-0.0993	0	79	0.00031	1	K.AFTSEEFTHFLEDLTK.Q

Proteins matching the same set of peptides:
gi|261342208 **Mass:** 45644 **Score:** 238 **Queries matched:** 4
phosphopyruvate hydratase [Enterobacter cancerogenus ATCC 35316]

Mascot Search Results

Protein View

Match to: **gi|261342208** Score: **238**
phosphopyruvate hydratase [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\280808\SSP 5702_RI12_01_475.d\SSP 5702_RI12_01_475.mgf

Nominal mass (M_r): **45644**; Calculated pI value: **5.19**
NCBI BLAST search of gi|261342208 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **12%**

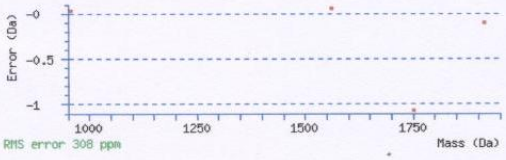
Matched peptides shown in **Bold Red**

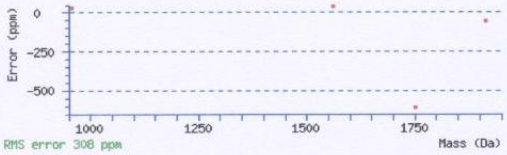
```
  1 MSKIVKVIGR EIIDSRGNPT VEAEVHLEGG FVGMAAAPSG ASTGSREALE
 51 LRDGDKSRFM GKGVLKAVGA VNGPIAQAIL GKDAKDQAGI DKIMIDLDGT
101 ENKSNFGANA ILAVSLANAK AAAAAKGQPL FEHIAELNGT PGKYSMPVPM
151 MNIINGGEHA DNNVDIQEFM IQPVGAKTLK EAVRMGSEVF HNLAKVLKAK
201 GMNTAVGDEG GYAPNLGSNA EALAVIAEAV KAAGYELGKD ITLAMDCAAS
251 EFYKDGKYVL AGEGNKAFTS EEFTHFLEDL TKQYPIVSIE DGLDESDWDG
301 FAYQTKVLGD KIQLVGDDLF VTNTKILKEG IEKGIVNSIL IKFNQIGSLT
351 ETLAAIKMAK DAGYTAVISH RSGETEDATI ADLAVGTAAG QIKTGSMSRS
401 DRVAKYNQLI RIEEALGEKA PYNGRKEIKG QA
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
240 - 254	875.8420	1749.6694	1750.7430	-1.0735	0	K.DITLAMDCAASEFYK.D Oxidation (M) (Ions score 22)
267 - 282	957.9100	1913.8054	1913.9047	-0.0993	0	K.AFTSEEFTHFLEDLTK.Q (Ions score 79)
312 - 325	781.9560	1561.8974	1561.8352	0.0623	0	K.IQLVGDDLFVTNTK.I (Ions score 92)
334 - 342	478.8270	955.6394	955.6066	0.0328	0	K.GIVNSILIK.F (Ions score 45)






```
LOCUS       ZP_05970066              432 aa            linear   BCT 15-OCT-2009
DEFINITION  phosphopyruvate hydratase [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05970066
VERSION     ZP_05970066.1  GI:261342208
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000033.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 432)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 432)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000033.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
```

4. gi|260771907 **Mass:** 45789 **Score:** 136 **Queries matched:** 2 **emPAI:** 0.07
enolase [Vibrio metschnikovii CIP 69.14]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
68	478.8270	955.6394	955.6066	0.0328	0	45	0.95	1	K.GIVNSILIK.F
117	781.9560	1561.8974	1561.8352	0.0623	0	92	2e-05	1	K.LQLVGDDLFVTNTK.I

5. gi|198241971 **Mass:** 45627 **Score:** 126 **Queries matched:** 2 **emPAI:** 0.07
phosphopyruvate hydratase [Salmonella enterica subsp. enterica serovar Dublin str. CT_02021853]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
117	781.9560	1561.8974	1561.8352	0.0623	0	92	2e-05	1	K.IQLVGDDLFVTNTK.I
☑ 148	875.8420	1749.6694	1749.7589	-0.0895	0	34	7.8	1	K.NITLAMDCAASEFYK.D + Oxidation (M)

6. gi|15642443 **Mass:** 45836 **Score:** 92 **Queries matched:** 1 **emPAI:** 0.07
phosphopyruvate hydratase [Vibrio cholerae O1 biovar El Tor str. N16961]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
117	781.9560	1561.8974	1561.8352	0.0623	0	92	2e-05	1	K.IQIVGDDLFVTNTK.I

Proteins matching the same set of peptides:
gi|29839249 **Mass:** 45848 **Score:** 92 **Queries matched:** 1
RecName: Full=Enolase; AltName: Full=2-phosphoglycerate dehydratase; AltName: Full=2-phospho-D-glycerate hydro-lyase
gi|77918834 **Mass:** 46350 **Score:** 92 **Queries matched:** 1
enolase [Pelobacter carbinolicus DSM 2380]
gi|90412096 **Mass:** 45596 **Score:** 92 **Queries matched:** 1
phosphopyruvate hydratase [Photobacterium profundum 3TCK]
gi|117617509 **Mass:** 45750 **Score:** 92 **Queries matched:** 1
phosphopyruvate hydratase [Aeromonas hydrophila subsp. hydrophila ATCC 7966]
gi|121591783 **Mass:** 33004 **Score:** 92 **Queries matched:** 1
enolase [Vibrio cholerae 2740-80]
gi|145300350 **Mass:** 45766 **Score:** 92 **Queries matched:** 1
phosphopyruvate hydratase [Aeromonas salmonicida subsp. salmonicida A449]
gi|153803630 **Mass:** 17966 **Score:** 92 **Queries matched:** 1
enolase, C- TIM barrel domain [Vibrio cholerae MZO-3]
gi|153827653 **Mass:** 40048 **Score:** 92 **Queries matched:** 1
enolase [Vibrio cholerae MZO-2]
gi|229527053 **Mass:** 45850 **Score:** 92 **Queries matched:** 1
enolase [Vibrio cholerae 12129(1)]
gi|237809481 **Mass:** 45583 **Score:** 92 **Queries matched:** 1
enolase [Tolumonas auensis DSM 9187]
gi|258626906 **Mass:** 45864 **Score:** 92 **Queries matched:** 1
Enolase [Vibrio mimicus VM603]
gi|261211542 **Mass:** 45820 **Score:** 92 **Queries matched:** 1
enolase [Vibrio sp. RC341]
gi|269123733 **Mass:** 46259 **Score:** 92 **Queries matched:** 1
enolase [Streptobacillus moniliformis DSM 12112]

7. gi|54310177 **Score:** 92 **Queries matched:** 1
phosphopyruvate hydratase [Photobacterium profundum SS9]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
117	781.9560	1561.8974	1561.8352	0.0623	0	92	2e-05	1	K.LQIVGDDLFVTNTK.I

Proteins matching the same set of peptides:
gi|84390114 **Score:** 92 **Queries matched:** 1
gi|86146349 **Score:** 92 **Queries matched:** 1
gi|218710558 **Score:** 92 **Queries matched:** 1
gi|260888598 **Score:** 92 **Queries matched:** 1

Peptide matches not assigned to protein hits: (no details means no match)

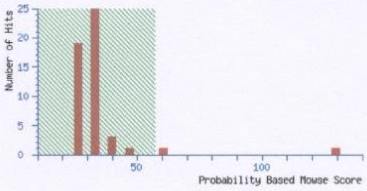
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 84	647.3270	1292.6394	1291.7136	0.9259	1	47	0.56	1	KTFADIGQTIAK
☑ 233	767.0460	2298.1162	2298.2947	-0.1785	1	39	3.2	1	QKPQNAQALETLDTIKFLLK
☑ 226	747.0360	2238.0862	2238.1314	-0.0453	1	29	27	1	SGLDGGGMITGSAGAFLVLESRK + Oxidation (M)
☑ 130	816.4800	2446.4182	2447.1421	-0.7239	2	29	55	1	DLTSIDHSTCECGRTLPRISK
☑ 209	1067.4960	3199.4662	3200.4945	-1.0284	2	27	73	1	TTTLQQRSTANVWDRFCEWITSTENR
☑ 100	714.6410	2140.9012	2140.1496	0.7515	2	27	75	1	LTMKGRAPDVLAMSATPIPR + Oxidation (M)
☑ 222	742.7000	2225.0782	2224.0206	1.0576	0	27	46	1	MTDVTATPASADPVAAYDPTSK + Oxidation (M)
☑ 158	902.4180	2704.2322	2704.3234	-0.0913	0	27	96	1	SSPDVGGQVLGMLPAGATLSLTGCMIR + Oxidation (M)
☑ 140	846.6630	2536.9672	2537.1793	-0.2121	2	27	71	1	TTFCVFLKAMFYGMEYSRGTK
☑ 51	428.2160	854.4174	855.5066	-1.0891	0	27	54	1	IATVLPDK
☑ 218	737.7080	2210.1022	2210.1219	-0.0197	1	26	54	1	VEAEKAFSTINTEFQNLAAK
☑ 93	686.3130	2055.9172	2057.0542	-1.1370	0	26	1.3e+02	1	VFTGNPALDAAAAPVTVSSNR
☑ 107	738.8360	2213.4862	2214.2115	-0.7254	1	26	1.3e+02	1	MSAVELLGINRDEVALIMIK
☑ 214	729.3010	2184.8812	2184.1397	0.7415	0	25	54	1	FILGFIMGAVLMCAIPSLAK + 2 Oxidation (M)
☑ 92	683.7950	1365.5754	1365.6268	-0.0514	0	25	1.4e+02	1	MTGSILVDDEMR
☑ 49	835.7180	834.7107	834.3905	0.3202	0	25	67	1	SLEGLCR

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 6203
MS data file    : D:\Data\Lakshmy\120808\241108\SSP 6203_RD22_01_943.d\SSP 6203_RD22_01_943.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 5 Jan 2010 at 11:47:36 GMT
Protein hits    : gi|129164 RecName: Full=Outer membrane protein X; Flags: Precursor
                  gi|440183 similar to outer membrane protein X from Enterobacter cloacae, Swiss-Prot Accession Number P25253; ORF2 [Es
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|129164 **Mass:** 18700 **Score:** 129 **Queries matched:** 2 **emPAI:** 0.39
RecName: Full=Outer membrane protein X; Flags: Precursor
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	78	708.7790	1415.5434	1415.6834	-0.1399	0	66	0.0056	1	K.GQYYGITAGPAYR.L
☑	118	898.9310	1795.8474	1796.9097	-1.0623	0	63	0.014	1	R.LNDWASIYGVVGVGYGK.F

Proteins matching the same set of peptides:
gi|16759751 **Mass:** 18541 **Score:** 129 **Queries matched:** 2
outer membrane protein X [Salmonella enterica subsp. enterica serovar Typhi str. CT18]
gi|152969400 **Mass:** 18452 **Score:** 129 **Queries matched:** 2
outer membrane protein X [Klebsiella pneumoniae subsp. pneumoniae MGH 78578]
gi|161504000 **Mass:** 18554 **Score:** 129 **Queries matched:** 2
outer membrane protein X [Salmonella enterica subsp. arizonae serovar 62:z4,z23:--]
gi|198245355 **Mass:** 18640 **Score:** 129 **Queries matched:** 2
outer membrane protein X [Salmonella enterica subsp. enterica serovar Dublin str. CT_02021853]
gi|205352088 **Mass:** 18540 **Score:** 129 **Queries matched:** 2
outer membrane protein X [Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91]
gi|206578835 **Mass:** 18528 **Score:** 129 **Queries matched:** 2
outer membrane protein X [Klebsiella pneumoniae 342]
gi|213583694 **Mass:** 14338 **Score:** 129 **Queries matched:** 2
outer membrane protein X [Salmonella enterica subsp. enterica serovar Typhi str. E98-0664]
gi|224582645 **Mass:** 18771 **Score:** 129 **Queries matched:** 2
outer membrane protein X [Salmonella enterica subsp. enterica serovar Paratyphi C strain RKS4594]
gi|261341400 **Mass:** 15248 **Score:** 129 **Queries matched:** 2
outer membrane protein X [Enterobacter cancerogenus ATCC 35316]

---

2. gi|440183 **Mass:** 14762 **Score:** 63 **Queries matched:** 1 **emPAI:** 0.23
similar to outer membrane protein X from Enterobacter cloacae, Swiss-Prot Accession Number P25253; ORF2 [Escherichia coli]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
118	898.9310	1795.8474	1796.9097	-1.0623	0	63	0.014	1	R.INDWASIYGVVGVGYGK.F

Proteins matching the same set of peptides:
gi|6435772 **Mass:** 16350 **Score:** 63 **Queries matched:** 1
Chain A, Crystal Structure Of The Outer Membrane Protein Ompx From Escherichia Coli
gi|15800566 **Mass:** 18649 **Score:** 63 **Queries matched:** 1
outer membrane protein X [Escherichia coli O157:H7 EDL933]
gi|26246790 **Mass:** 18879 **Score:** 63 **Queries matched:** 1
outer membrane protein X [Escherichia coli CFT073]
gi|30749947 **Mass:** 16390 **Score:** 63 **Queries matched:** 1
Chain A, Nmr Fold Of The Outer Membrane Protein Ompx In Dhpc Micelles

```
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..432
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..432
                     /product="phosphopyruvate hydratase"
                     /EC_number="4.2.1.11"
                     /calculated_mol_wt=45484
     Region          1..432
                     /region_name="eno"
                     /note="phosphopyruvate hydratase; Provisional; PRK00077"
                     /db_xref="CDD:134085"
     Region          6..414
                     /region_name="enolase"
                     /note="Enolase: Enolases are homodimeric enzymes that
                     catalyse the reversible dehydration of
                     2-phospho-D-glycerate to phosphoenolpyruvate as part of
                     the glycolytic and gluconeogenesis pathways. The reaction
                     is facilitated by the presence of metal ions; cd03313"
                     /db_xref="CDD:48188"
     Site            order(8,10..18,22,34,160..162,183..184,187..188,191..192,
                     205..206,213,215,374..376,399..401,403..404,407,410..411,
                     414)
                     /site_type="other"
                     /note="dimer interface"
                     /db_xref="CDD:48188"
     Site            order(42,246,290,317)
                     /site_type="metal-binding"
                     /db_xref="CDD:48188"
     Site            order(159,209,342,370..372,393)
                     /site_type="other"
                     /note="substrate binding pocket"
                     /db_xref="CDD:48188"
     CDS             1..432
                     /gene="eno"
                     /locus_tag="EcanA3_020100017202"
                     /coded_by="complement(NZ_ABWM02000033.1:143450..144748)"
                     /note="COG0148 Enolase"
                     /transl_table=11
                     /db_xref="CDD:134085"
```

Mascot: http://www.matrixscience.com/

Mascot Search Results

Protein View

Match to: **gi|261341400** Score: **129**
outer membrane protein X [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\241108\SSP 6203_RD22_01_943.d\SSP 6203_RD22_01_943.mgf

Nominal mass (M_r): **15248**; Calculated pI value: **5.38**
NCBI BLAST search of gi|261341400 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **22%**

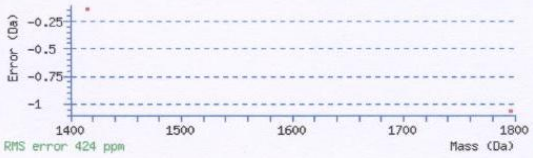
Matched peptides shown in **Bold Red**

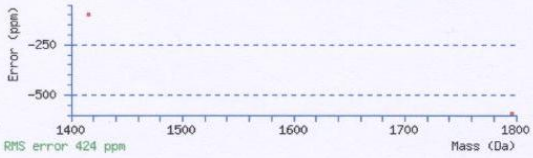
```
  1 MQGVMNKTNG FNLKYRYEQD NNPLGVIGSF TYTEKDRTEN GAYNKGQYYG
 51 ITAGPAYRLN DWASIYGVVG VGYGKFQQTE NEGLNRTASN SDYGFSYGAG
101 MQFNPIENVA LDFSYEQSRI RNVDVGTWIA GVGYRF
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
46 - 58	708.7790	1415.5434	1415.6834	-0.1399	0	K.GQYYGITAGPAYR.L (Ions score 66)
59 - 75	898.9310	1795.8474	1796.9097	-1.0623	0	R.LNDWASIYGVVGVGYGK.F (Ions score 63)






```
LOCUS       ZP_05969258              136 aa            linear   BCT 15-OCT-2009
DEFINITION  outer membrane protein X [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05969258
VERSION     ZP_05969258.1  GI:261341400
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000020.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 136)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 136)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000020.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
```

```
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..136
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..136
                     /product="outer membrane protein X"
                     /calculated_mol_wt=15127
     Region          <1..136
                     /region_name="Surface_Ag_2"
                     /note="Surface antigen; cl01155"
                     /db_xref="CDD:141128"
     CDS             1..136
                     /gene="ompX"
                     /locus_tag="EcanA3_020100013096"
                     /coded_by="NZ_ABWM02000020.1:323158..323568"
                     /note="COG3637 Opacity protein and related surface
                     antigens"
                     /transl_table=11
                     /db_xref="CDD:137256"
```

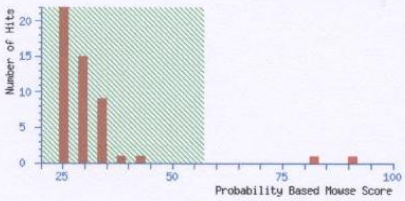
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 6306
MS data file    : D:\Data\Lakshmy\120808\SSP 6306_RB6_01_719.d\SSP 6306_RB6_01_719.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 5 Jan 2010 at 11:53:30 GMT
Protein hits    : gi|152972668 hypothetical protein KPN_04168 [Klebsiella pneumoniae subsp. pneumoniae MGH 78578]
                  gi|50582463  oxidoreductase [Exiguobacterium acetylicum]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|152972668 **Mass:** 10606 **Score:** 91 **Queries matched:** 3 **emPAI:** 0.75
hypothetical protein KPN_04168 [Klebsiella pneumoniae subsp. pneumoniae MGH 78578]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	104	753.3170	1504.6194	1504.7595	-0.1401	0	48	0.43	1	K.EPELNLMQFLQK.L + Oxidation (M)
☑	123	840.9220	1679.8294	1679.9246	-0.0952	0	43	1.3	1	R.LNEVIELLQPAWQK.E
	124	560.9530	1679.8372	1679.9246	-0.0874	0	(29)	36	3	R.LNEVIELLQPAWQK.E

Proteins matching the same set of peptides:
gi|206579500 **Mass:** 10388 **Score:** 91 **Queries matched:** 3
hypothetical protein KPK_5514 [Klebsiella pneumoniae 342]
gi|261342932 **Mass:** 10363 **Score:** 91 **Queries matched:** 3
hypothetical protein EcanA3_20877 [Enterobacter cancerogenus ATCC 35316]

2. gi|50582463 **Mass:** 27360 **Score:** 83 **Queries matched:** 3 **emPAI:** 0.12
oxidoreductase [Exiguobacterium acetylicum]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	174	677.2520	2028.7342	2028.9752	-0.2411	0	24	67	9	K.SVDLADNQNVHDLYEGLK.E
☑	250	984.4460	2950.3162	2950.4236	-0.1075	0	58	0.025	1	K.ELDIETWINNAGFGDFDLVQDIELGK.I
☑	162	984.7600	2951.2582	2950.4236	0.8345	0	(55)	0.096	1	K.ELDIETWINNAGFGDFDLVQDIELGK.I

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	85	610.2780	1218.5414	1218.6972	-0.1558	0	41	2.7	1	LTDTQLFLLR
☑	149	948.4010	1894.7874	1894.9312	-0.1438	0	39	2.9	1	IGDTAPAQYQTPEYTIK
☑	220	761.6840	2282.0302	2282.2867	-0.2565	0	35	6.5	1	LLEYGLGIGYLLLFIPFWR
☑	124	560.9530	1679.8372	1680.0371	-0.2000	2	35	8.2	1	KLELVPMIKLIIDR
☑	75	563.2460	1686.7162	1686.9212	-0.2051	2	34	21	1	VHHGVARALLAACRR
☑	234	837.2810	2508.8212	2509.1802	-0.3590	2	33	8	1	DGARIGTTGNGIGPCYADLAARMR + Oxidation (M)
☑	82	602.2490	1202.4834	1202.6441	-0.1607	1	32	19	1	LTDAQVKAAMR
☑	88	629.2630	1256.5114	1255.6996	0.8118	1	32	18	1	AEALRQGATLAR
☑	248	972.1250	2913.3532	2912.4425	0.9107	1	30	16	1	VYRNHINNGVEMIMAGHIALPEYQK + Oxidation (M)
☑	73	562.4190	1122.8234	1121.6346	1.1889	2	30	20	1	FQRFLQGKV
☑	157	647.5700	1939.6882	1939.8689	-0.1808	0	29	20	1	MGIAEGDMIGTAAGLATCGK + Oxidation (M)
☑	175	677.2520	2028.7342	2027.9881	0.7461	1	29	23	1	AVADEGARGVLPAGPMQAMR + 2 Oxidation (M)
☑	174	677.2520	2028.7342	2028.8881	-0.1540	0	29	24	1	GSGMTTMDHSFVGSTALGNK + 2 Oxidation (M)
☑	222	767.0440	2298.1102	2298.2107	-0.1006	1	28	34	1	FYDVSSIGLGDLGQGTLTKLSK
☑	118	822.2250	2463.6532	2464.2420	-0.5888	0	28	49	1	YAAESFTPNRPVLEAALELMAR + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|261342932** Score: **91**
hypothetical protein EcanA3_20877 [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\SSP 6306_RB6_01_719.d\SSP 6306_RB6_01_719.mgf

Nominal mass (M_r): **10363**; Calculated pI value: **5.02**
NCBI BLAST search of gi|261342932 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **29%**

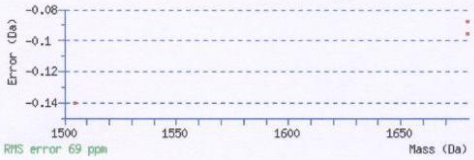
Matched peptides shown in **Bold Red**

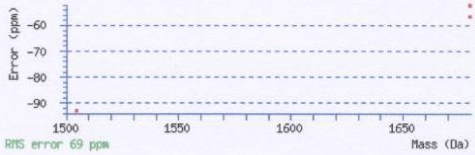
1 MKCKR**LNEVI ELLQPAWQKE PELNLMQFLQ K**LAKESGFDG ELTDLSDDIL
51 IYQLKMRDSA KDAVIPGIQK DYEEDFKTAL LRARGVIKE

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence	
6 - 19	840.9220	1679.8294	1679.9246	-0.0952	0	R.LNEVIELLQPAWQK.E	(Ions score 43)
6 - 19	560.9530	1679.8372	1679.9246	-0.0874	0	R.LNEVIELLQPAWQK.E	(Ions score 29)
20 - 31	753.3170	1504.6194	1504.7595	-0.1401	0	K.EPELNLMQFLQK.L Oxidation (M)	(Ions score 48)






```
LOCUS       ZP_05970790                89 aa            linear   BCT 15-OCT-2009
DEFINITION  hypothetical protein EcanA3_20877 [Enterobacter cancerogenus ATCC
            35316].
ACCESSION   ZP_05970790
VERSION     ZP_05970790.1  GI:261342932
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000045.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 89)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 89)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000045.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
```

```
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..89
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..89
                     /product="hypothetical protein"
                     /calculated_mol_wt=10181
     Region          1..86
                     /region_name="DUF1040"
                     /note="Protein of unknown function (DUF1040); cl01176"
                     /db_xref="CDD:120451"
     CDS             1..89
                     /locus_tag="EcanA3_020100020877"
                     /coded_by="NZ_ABWM02000045.1:1664..1933"
                     /note="COG3084 Uncharacterized protein conserved in
                     bacteria"
                     /transl_table=11
                     /db_xref="CDD:114979"
```

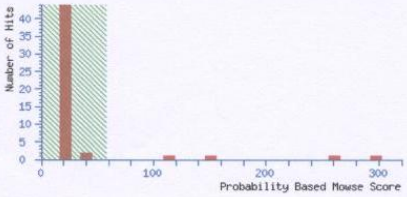
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 6401
MS data file    : D:\Data\Lakshmy\120808\280808\SSP 6401_RK4_01_560.d\SSP 6401_RK4_01_560.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 5 Jan 2010 at 11:56:40 GMT
Protein hits    : gi|195940647 cytidylate kinase [Escherichia coli O157:H7 str. EC4024]
                  gi|6137462   Chain A, Cmp Kinase From Escherichia Coli Free Enzyme Structure
                  gi|152969494 cytidylate kinase [Klebsiella pneumoniae subsp. pneumoniae MGH 78578]
                  gi|261339244 cytidylate kinase [Enterobacter cancerogenus ATCC 35316]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 58 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|195940647 **Mass:** 24783 **Score:** 297 **Queries matched:** 5 **emPAI:** 0.46
cytidylate kinase [Escherichia coli O157:H7 str. EC4024]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	130	696.3070	1390.5994	1390.6803	-0.0808	0	68	0.0041	1	R.DMGTVVFPDAPVK.I + Oxidation (M)
☑	150	784.4830	1566.9514	1566.8617	0.0897	0	79	0.00031	1	M.TAIAPVITIDGPSGAGK.G
☑	161	830.4350	1658.8554	1658.8376	0.0178	0	90	2.9e-05	1	R.TQEVANAASQVAAFPR.V
☑	247	793.4150	2377.2232	2377.1649	0.0583	0	40	2.4	1	R.FVSTDGNLEVILEGEDVSGEIR.T
	250	940.8840	2819.6302	2820.5372	-0.9070	0	20	1.5e+02	3	R.AVAPLVPAEDALVLDSTSLTIEQVIEK.A

2. gi|6137462 **Mass:** 24759 **Score:** 264 **Queries matched:** 4 **emPAI:** 0.46
Chain A, Cmp Kinase From Escherichia Coli Free Enzyme Structure
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
130	696.3070	1390.5994	1390.6803	-0.0808	0	68	0.0041	1	R.DMGTVVFPDAPVK.I + Oxidation (M)
150	784.4830	1566.9514	1566.8617	0.0897	0	79	0.00031	1	M.TAIAPVITIDGPSGAGK.G
161	830.4350	1658.8554	1658.8376	0.0178	0	90	2.9e-05	1	R.TQEVANAASQVAAFPR.V
247	793.4150	2377.2232	2376.1809	1.0423	0	27	49	3	R.FVSTNGNLEVILEGEDVSGEIR.T

Proteins matching the same set of peptides:
gi|15800771 **Mass:** 24789 **Score:** 264 **Queries matched:** 4
cytidylate kinase [Escherichia coli O157:H7 EDL933]
gi|24112319 **Mass:** 24803 **Score:** 264 **Queries matched:** 4
cytidylate kinase [Shigella flexneri 2a str. 301]
gi|74311467 **Mass:** 24805 **Score:** 264 **Queries matched:** 4
cytidylate kinase [Shigella sonnei Ss046]
gi|82777567 **Mass:** 24729 **Score:** 264 **Queries matched:** 4
cytidylate kinase [Shigella dysenteriae Sd197]
gi|88192972 **Mass:** 24764 **Score:** 264 **Queries matched:** 4
Chain A, Mutant R188m Of The Cytidine Monophosphate Kinase From E. Coli
gi|157146404 **Mass:** 24818 **Score:** 264 **Queries matched:** 4
cytidylate kinase [Citrobacter koseri ATCC BAA-895]

3. gi|152969494 **Mass:** 24917 **Score:** 158 **Queries matched:** 2 **emPAI:** 0.29
cytidylate kinase [Klebsiella pneumoniae subsp. pneumoniae MGH 78578]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
130	696.3070	1390.5994	1390.6803	-0.0808	0	68	0.0041	1	R.DMGTVVFPDAPVK.I + Oxidation (M)
161	830.4350	1658.8554	1658.8740	-0.0185	1	90	2.9e-05	1	R.TQEVANAASKVAAFPR.V

Proteins matching the same set of peptides:
gi|206579846 **Mass:** 24874 **Score:** 158 **Queries matched:** 2
cytidylate kinase [Klebsiella pneumoniae 342]
gi|238893966 **Mass:** 24931 **Score:** 158 **Queries matched:** 2
cytidylate kinase [Klebsiella pneumoniae NTUH-K2044]
gi|262040991 **Mass:** 27105 **Score:** 156 **Queries matched:** 2
cytidylate kinase [Klebsiella pneumoniae subsp. rhinoscleromatis ATCC 13884]

---

4. gi|261339244 **Mass:** 24768 **Score:** 115 **Queries matched:** 3 **emPAI:** 0.14
cytidylate kinase [Enterobacter cancerogenus ATCC 35316]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
130	696.3070	1390.5994	1390.6803	-0.0808	0	68	0.0041	1	R.DMGTVVFPDAPVK.I + Oxidation (M)
247	793.4150	2377.2232	2376.1809	1.0423	0	27	49	3	R.FVSTNGNLEVILEGEDVSGEIR.T
250	940.8840	2819.6302	2820.5372	-0.9070	0	20	1.5e+02	3	R.AVAPLVPAEDALVLDSTSLTIEQVIEK.A

Peptide matches not assigned to protein hits: (no details means no match)

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
198	971.0240	1940.0334	1939.8266	0.2068	1	46	0.71	1	MNFEKFGEMYLQSCR
238	737.7190	2210.1352	2210.1439	-0.0087	1	30	25	1	DAYLIIGGGTIISMQAGMKGAK + Oxidation (M)
223	1043.4990	2084.9834	2084.0432	0.9402	1	30	45	1	SERSNGAGAVLMHALESAIR + Oxidation (M)
106	565.3480	1693.0222	1693.8471	-0.8249	2	30	63	1	EMTRSFQLGRWQR
243	747.0480	2238.1222	2238.0409	0.0813	1	28	36	1	LEQMTRSEASQWLEMQGAK + Oxidation (M)
189	623.0330	1866.0772	1865.7148	0.3624	0	28	40	1	QDGAGESDGSGDGAESGDGAK
142	759.4010	2275.1812	2276.0295	-0.8484	0	28	90	1	MMNWYWQAWQLGLSFSGR + Oxidation (M)
241	742.3940	2224.1602	2223.0240	1.1362	2	26	61	1	MPRGGAFAWGGSRTQTDAEGR + Oxidation (M)
144	771.9940	2312.9602	2313.0954	-0.1352	2	26	1.3e+02	1	GGRDEVMARLHQELVACAER + Oxidation (M)
197	647.6310	1939.8712	1939.9356	-0.0644	1	25	84	1	MLEQAIHQRDAGACLAR
120	642.3800	1924.1182	1924.0088	0.1094	2	24	2e+02	1	VKGTTMVAQTGGDGKYAIK
250	940.8840	2819.6302	2819.5197	0.1105	1	24	64	1	TRASTIMHGFWILLFAVPFAGLVEK + Oxidation (M)
199	972.5850	2914.7332	2914.4297	0.3035	2	24	1.6e+02	1	VHPYRLNAFASGEAGPLGWMEEGRVR + Oxidation (M)
183	910.0500	2727.1282	2726.3407	0.7875	1	24	1.8e+02	1	LMLQKAATLYDAGQEFAAAEAANMAK
205	1002.0500	2002.0854	2000.9480	1.1375	0	23	1.2e+02	1	LITWSSDYDDGFSVALGR
93	491.3570	1471.0492	1470.8042	0.2450	1	23	2.1e+02	1	AVIAREGIEASVEK
173	879.1510	2634.4312	2634.4315	-0.0004	2	23	1.9e+02	1	YAAMPHLKKAELIEGVVHLSSAVR + Oxidation (M)
230	1072.9460	2143.8774	2143.1830	0.6944	2	23	1e+02	1	KLYVSLVKNGYPFVFVDR
195	645.3600	1933.0582	1932.9615	0.0967	0	23	1.4e+02	1	DITTLAAAMAAIAPDFDAR
111	587.8760	1760.6062	1760.9156	-0.3094	1	23	2.7e+02	1	VTVNESDLAALGEKSTK
126	676.0180	2025.0322	2025.8846	-0.8524	0	23	2.4e+02	1	SEYSGVHEVMAMPMGIEK + 2 Oxidation (M)
129	691.9220	1381.8294	1380.7222	1.1072	1	22	1.6e+02	1	GVRGGAGGPAGLAEGR
118	632.3890	1894.1452	1894.9981	-0.8530	2	22	3e+02	1	QAKGVAALMRMLANEHR
196	970.7840	2909.3302	2910.4592	-1.1291	2	22	2e+02	1	LNQGMRLQTDPTVIYGMGSAYNGRIR
202	982.2250	2943.6532	2944.4569	-0.8037	1	22	2.1e+02	1	MSAMDAGIPAGTTVAAVMQRAPVAQLEAR + 2 Oxidation (M)
48	375.1880	1122.5422	1123.6349	-1.0928	0	22	3.6e+02	1	VQALLEAAPGR
190	936.7680	2807.2822	2806.3068	0.9754	2	22	2.3e+02	1	FGHRYGGCGSPLRVPGYGDVPAAVCR
240	1110.5770	3328.7092	3327.7615	0.9477	2	22	2.5e+02	1	GEVRPYYQAVGSEEAVFKAAYRQGLSLVLK
137	735.8300	2204.4682	2203.9700	0.4981	1	22	2.8e+02	1	GQHSICCVMTEAFLRYSK + Oxidation (M)
248	853.9570	2558.8492	2559.3526	-0.5034	2	22	98	1	VGGATYQVPVEVRPARRMALAMR + 2 Oxidation (M)
231	1075.6540	3223.9402	3223.6005	0.3396	2	22	2.3e+02	1	VLWKQNIYAESGMGCTGPIIRVSDANLSK + Oxidation (M)
222	1042.1070	2082.1994	2083.0334	-0.8340	2	22	1.6e+02	1	SGYERVDIVENKGEFSVR
155	811.2030	2430.5872	2431.2529	-0.6657	2	21	2.5e+02	1	NPLMIIGGFARSLLKNSDENDK
128	680.9010	1359.7874	1359.6895	0.0980	0	21	3.4e+02	1	EALEAEGHIHVR
184	911.8940	2732.6602	2732.3922	0.2680	2	21	2.9e+02	1	LREFFSSRSYSPTIIGYTNELPR
162	838.7720	2513.2942	2514.1155	-0.8213	1	21	3.2e+02	1	SPMFNGNIQSTGPRYCPSIEDK + Oxidation (M)
53	391.1710	1170.4912	1171.5330	-1.0418	1	21	4.2e+02	1	AQQGNGPSSRGN
194	964.0440	2889.1102	2889.3744	-0.2643	1	21	3.2e+02	1	MKLTQLFASLMVAGALTACAQNGADMK + 3 Oxidation (M)
166	844.0780	2529.2122	2530.3465	-1.1344	1	21	3.4e+02	1	LGAAGVMVIGDNLQVVFGPKSDSIK + Oxidation (M)
98	514.3690	1540.0852	1539.7682	0.3170	0	21	3.9e+02	1	VPADVASPWLSDAGR
171	860.3590	2578.0552	2577.3360	0.7192	1	21	3.4e+02	1	QPELVEVLDTHLLPQDKDTIMK + Oxidation (M)
233	1082.6550	3244.9432	3245.6179	-0.6747	2	21	2.8e+02	1	EHQFEFKQGPIFTQILLADEINRCSPK
102	553.9120	1105.8094	1106.5179	-0.7084	0	21	2e+02	1	GYNIVACGPR
104	557.3730	1669.0972	1668.7930	0.3042	1	21	4.4e+02	1	HSGYMVGGTSPFGTKK + Oxidation (M)
179	902.2060	2703.5962	2703.3438	0.2523	0	20	3.1e+02	1	LAALMQNLGEYNVQLEEIHHTHK + Oxidation (M)
168	849.9520	2546.8342	2546.3162	0.5179	1	20	3.8e+02	1	ARSASPEVVLITAYGTPAAAVEAMR + Oxidation (M)
145	772.8310	2315.4712	2316.1420	-0.6708	1	20	3.9e+02	1	DEETPVNTIPIDAMFSPIRR + Oxidation (M)
226	1047.7250	3140.1532	3140.6401	-0.4869	1	20	2.6e+02	1	AASSAGVLHGVSGLFMAGITIAKGANGAGALTDR
135	479.2060	1434.5962	1435.7459	-1.1498	0	20	2.4e+02	1	AWLPIDPATPAER
151	790.3940	2368.1602	2369.1222	-0.9621	0	20	4.8e+02	1	NVNILMPAEYAPHHNSYIDR + Oxidation (M)
224	1046.0650	3135.1732	3135.6434	-0.4702	1	20	3.8e+02	1	GGHASMPHAAADPVFVAAEIVLALQAMVTRR
158	824.1480	2469.4222	2470.1917	-0.7695	2	20	4e+02	1	DKDGGPLIPWDGPESAFEAWKR
175	895.7880	2684.3422	2683.2661	1.0761	1	20	4.1e+02	1	GQVCGLKWSDVDLDAGQITVHDNR

```
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..227
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..227
                     /product="cytidylate kinase"
                     /calculated_mol_wt=24594
     Region          5..225
                     /region_name="cmk"
                     /note="cytidylate kinase; Provisional; PRK00023"
                     /db_xref="CDD:134038"
     Region          7..205
                     /region_name="CMPK"
                     /note="Cytidine monophosphate kinase (CMPK) catalyzes the
                     reversible phosphorylation of cytidine monophosphate (CMP)
                     to produce cytidine diphosphate (CDP), using ATP as the
                     preferred phosphoryl donor; cd02020"
                     /db_xref="CDD:73296"
     Site            order(110,131..132,188)
                     /site_type="other"
                     /note="CMP-binding site"
                     /db_xref="CDD:73296"
     Site            order(181,185)
                     /site_type="other"
                     /note="The sites determining sugar specificity"
                     /db_xref="CDD:73296"
     CDS             1..227
                     /locus_tag="EcanA3_020100002225"
                     /coded_by="NZ_ABWM02000004.1:495167..495850"
                     /note="COG0283 Cytidylate kinase"
                     /transl_table=11
                     /db_xref="CDD:134038"
```

Mascot: http://www.matrixscience.com/

Mascot Search Results

Protein View

Match to: **gi|261339244** Score: **115**
cytidylate kinase [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\280808\SSP 6401_RK4_01_560.d\SSP 6401_RK4_01_560.mgf

Nominal mass (M_r): **24768**; Calculated pI value: **5.32**
NCBI BLAST search of gi|261339244 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **27%**

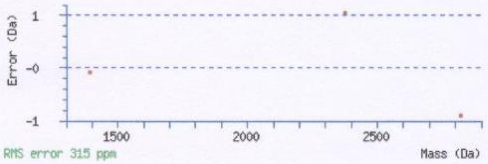
Matched peptides shown in **Bold Red**

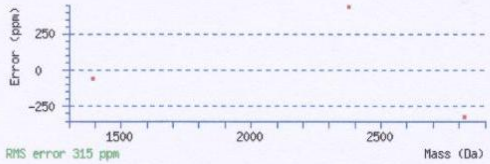
```
  1 MTAVAPVITI DGPSGAGKGT LCKAMAEALQ WHLLDSGAIY RVLALAALHH
 51 HVDVASEEAL VPLAAHLDVR FVSTNGNLEV ILEGEDVSGE IRTQDVANAA
101 SQVAAFPRVR EALLRRQRGF REAPGLIADG RDMGTVVFPD APVKIFLDAS
151 SEERAQRRML QLQEKGFSVN FERLLSEIKE RDDRDRNRAV APLVPAEDAL
201 VLDSTSLTIE QVIEKALQYA RQKLALA
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence	
71 - 92	793.4150	2377.2232	2376.1809	1.0423	0	R.FVSTNGNLEVILEGEDVSGEIR.T	(Ions score 27)
132 - 144	696.3070	1390.5994	1390.6803	-0.0808	0	R.DMGTVVFPDAPVK.I Oxidation (M)	(Ions score 68)
189 - 215	940.8840	2819.6302	2820.5372	-0.9070	0	R.AVAPLVPAEDALVLDSTSLTIEQVIEK.A	(Ions score 20)






```
LOCUS       ZP_05967102              227 aa            linear   BCT 15-OCT-2009
DEFINITION  cytidylate kinase [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05967102
VERSION     ZP_05967102.1  GI:261339244
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000004.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 227)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 227)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000004.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
```

```
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..227
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..227
                     /product="cytidylate kinase"
                     /calculated_mol_wt=24594
     Region          5..225
                     /region_name="cmk"
                     /note="cytidylate kinase; Provisional; PRK00023"
                     /db_xref="CDD:134038"
     Region          7..205
                     /region_name="CMPK"
                     /note="Cytidine monophosphate kinase (CMPK) catalyzes the
                     reversible phosphorylation of cytidine monophosphate (CMP)
                     to produce cytidine diphosphate (CDP), using ATP as the
                     preferred phosphoryl donor; cd02020"
                     /db_xref="CDD:73296"
     Site            order(110,131..132,188)
                     /site_type="other"
                     /note="CMP-binding site"
                     /db_xref="CDD:73296"
     Site            order(181,185)
                     /site_type="other"
                     /note="The sites determining sugar specificity"
                     /db_xref="CDD:73296"
     CDS             1..227
                     /locus_tag="EcanA3_020100002225"
                     /coded_by="NZ_ABWM02000004.1:495167..495850"
                     /note="COG0283 Cytidylate kinase"
                     /transl_table=11
                     /db_xref="CDD:134038"
```

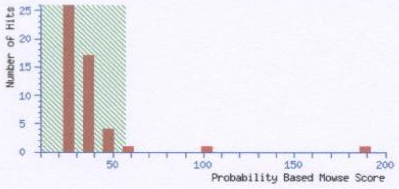
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User          : Lakshmy Manickan
Email         : lakshmy.manickan@unn.ac.uk
Search title  : SSP 6408
MS data file  : D:\Data\Lakshmy\120808\280808\SSP 6408_RI7_01_467.d\SSP 6408_RI7_01_467.mgf
Database      : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy      : Bacteria (Eubacteria) (5690016 sequences)
Timestamp     : 5 Jan 2010 at 11:58:13 GMT
Protein hits  : gi|261342934 periplasmic protein disulfide isomerase I [Enterobacter cancerogenus ATCC 35316]
                gi|2558846   disulfide oxidoreductase [Salmonella enterica subsp. enterica serovar Typhimurium]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|261342934 **Mass:** 23134 **Score:** 189 **Queries matched:** 5 **emPAI:** 0.50
periplasmic protein disulfide isomerase I [Enterobacter cancerogenus ATCC 35316]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
73	630.3790	1258.7434	1258.6710	0.0724	0	42	2.1	5	K.YHVEFLGPLGK.D
☑ 114	800.8990	1599.7834	1599.7205	0.0629	0	67	0.006	1	K.GEDYDAAWNSFVVK.S
☑ 155	600.6850	1799.0332	1798.9465	0.0867	0	(44)	1.1	1	K.DLTQAWAVAIALGVEDK.V
☑ 156	900.5290	1799.0434	1798.9465	0.0969	0	80	0.00024	1	K.DLTQAWAVAIALGVEDK.V
☑ 157	900.5730	1799.1314	1798.9465	0.1849	0	(50)	0.21	1	K.DLTQAWAVAIALGVEDK.V

Proteins matching the same set of peptides:
gi|152972670 **Mass:** 23047 **Score:** 187 **Queries matched:** 5
periplasmic protein disulfide isomerase I [Klebsiella pneumoniae subsp. pneumoniae MGH 78578]
gi|206577959 **Mass:** 23034 **Score:** 187 **Queries matched:** 5
thiol:disulfide interchange protein DsbA [Klebsiella pneumoniae 342]

---

2. gi|2558846 **Score:** 102 **Queries matched:** 2
disulfide oxidoreductase [Salmonella enterica subsp. enterica serovar Typhimurium]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
73	630.3790	1258.7434	1258.6710	0.0724	0	42	2.1	5	K.YHVEFLGPLGK.E
114	800.8990	1599.7834	1598.7365	1.0469	0	61	0.025	2	K.GENYDAAWNSFVVK.S

Peptide matches not assigned to protein hits: (no details means no match)

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 70	608.8640	1215.7134	1215.6870	0.0265	2	53	0.16	1	LTAPLMRRSR + Oxidation (M)
☑ 73	630.3790	1258.7434	1258.6128	0.1306	1	51	0.25	1	RAWGCLDVPGK
☑ 72	420.5380	1258.5922	1257.6717	0.9204	0	35	11	1	ALLDGLFAGPER
☑ 219	747.0790	2238.2152	2238.1202	0.0950	1	34	8.8	1	THMVTKSYVLGEGETVSSVAK + Oxidation (M)
☑ 197	1042.0290	2082.0434	2082.0456	-0.0021	1	32	14	1	VMNKFVDIELVDGIGDFR + Oxidation (M)
☑ 226	767.4120	2299.2142	2298.2947	0.9195	1	32	14	1	QKPQNAQALETLDTIKFLLK
☑ 223	762.1160	2283.3262	2284.0971	-0.7710	1	31	18	1	KWTIENQGSFLNSDATASTSK
☑ 212	1106.1540	2210.2934	2210.0598	0.2336	2	31	19	1	YNQAKMPEIELVDMKEEK + Oxidation (M)
☑ 76	685.7070	2054.0992	2054.1850	-0.0858	1	29	59	1	VFMLSSAVLVSLYILKQK + Oxidation (M)
☑ 154	897.9300	1793.8454	1793.8692	-0.0237	1	28	38	1	MKILTSYAYGDIMTR + 2 Oxidation (M)
☑ 125	828.9490	2483.8252	2484.3019	-0.4768	2	28	74	1	MQRSRPPFRADEVGSLLRPEK + Oxidation (M)
☑ 189	1007.7040	3020.0902	3019.4855	0.6047	2	27	55	1	AGLADRVGTLDLAIAEMAADFDRMASAPR + Oxidation (M)
☑ 48	435.7510	869.4874	870.4923	-1.0049	1	27	53	1	KDGVTVPR

Mascot Search Results

Protein View

Match to: **gi|261342934** Score: **189**
periplasmic protein disulfide isomerase I [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\280808\SSP 6408_RI7_01_467.d\SSP 6408_RI7_01_467.mgf

Nominal mass (M_r): **23134**; Calculated pI value: **5.65**
NCBI BLAST search of gi|261342934 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **20%**

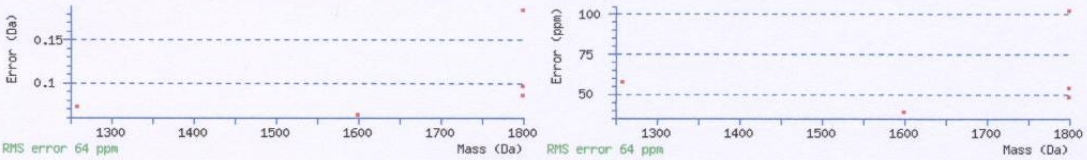
Matched peptides shown in **Bold Red**

```
  1 MKKIWLALAG MILAFSASAA QFTDGKQYIT LDKPVAGEPQ VLEFFSFYCP
 51 HCYQFEEVLH VSDNVKKKLP EGTKMTKYHV EFLGPLGKDL TQAWAVAIAL
101 GVEDKVTAPL FEAVQKTQTV QNTADIRKVF VDAGVKGEDY DAAWNSFVVK
151 SLVAQQEKAA ADLQLQGVPA MFVNGKYQLN MQGMDTSSMD IFVQQYADTV
201 KYLVEKK
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
78 - 88	630.3790	1258.7434	1258.6710	0.0724	0	K.YHVEFLGPLGK.D (Ions score 42)
89 - 105	600.6850	1799.0332	1798.9465	0.0867	0	K.DLTQAWAVAIALGVEDK.V (Ions score 44)
89 - 105	900.5290	1799.0434	1798.9465	0.0969	0	K.DLTQAWAVAIALGVEDK.V (Ions score 80)
89 - 105	900.5730	1799.1314	1798.9465	0.1849	0	K.DLTQAWAVAIALGVEDK.V (Ions score 50)
137 - 150	800.8990	1599.7834	1599.7205	0.0629	0	K.GEDYDAAWNSFVVK.S (Ions score 67)




```
LOCUS       ZP_05970792              207 aa            linear   BCT 15-OCT-2009
DEFINITION  periplasmic protein disulfide isomerase I [Enterobacter
            cancerogenus ATCC 35316].
ACCESSION   ZP_05970792
VERSION     ZP_05970792.1  GI:261342934
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000045.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 207)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 207)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000045.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
```

found here:
http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be aware that the annotation is done automatically with little or no manual curation.
Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter cancerogenus, not the sequenced strain), is a member of the Proteobacteria division of the domain bacteria and has been isolated from human feces. The sequenced strain was obtained from ATCC (ATCC 35316).

We have performed one round of automated sequence improvement (pre-finishing), along with manual improvement that includes breaking apart any mis-assembly, and making manual joins where possible. Manual edits also are made where the consensus appears to be incorrect. All low quality data on the ends of contigs is removed. Contigs are ordered and oriented where possible.

This work was supported by the National Human Genome Research Institute and the NIH Roadmap Human Microbiome Project [http://nihroadmap.nih.gov/hmp/].

This is a reference genome for the Human Microbiome Project.
This project is co-owned with the Human Microbiome Project DACC.
Genome Coverage: 29.7x
Sequencing Technology: 454.
Method: conceptual translation.

```
FEATURES             Location/Qualifiers
     source          1..207
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..207
                     /product="periplasmic protein disulfide isomerase I"
                     /calculated_mol_wt=22901
     Region          24..202
                     /region_name="DsbA_DsbA"
                     /note="DsbA family, DsbA subfamily; DsbA is a monomeric
                     thiol disulfide oxidoreductase protein containing a redox
                     active CXXC motif imbedded in a TRX fold. It is involved
                     in the oxidative protein folding pathway in prokaryotes,
                     and is the strongest thiol...; cd03019"
                     /db_xref="CDD:48568"
     Region          40..197
                     /region_name="DSBA"
                     /note="DSBA-like thioredoxin domain; pfam01323"
                     /db_xref="CDD:110334"
     Site            order(49,51..52,169)
                     /site_type="other"
                     /note="catalytic residues"
                     /db_xref="CDD:48568"
     Site            82..84
                     /site_type="other"
                     /note="hinge region"
                     /db_xref="CDD:48568"
     Site            order(89..97,102..121,124..132,138..147)
                     /site_type="other"
                     /note="alpha helical domain"
                     /db_xref="CDD:48568"
     CDS             1..207
                     /locus_tag="EcanA3_020100020887"
                     /coded_by="NZ_ABWM02000045.1:3023..3646"
                     /note="COG0526 Thiol-disulfide isomerase and thioredoxins"
                     /transl_table=11
                     /db_xref="CDD:138304"
```

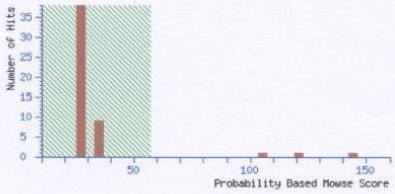
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User           : Lakshmy Manickan
Email          : lakshmy.manickan@unn.ac.uk
Search title   : SSP 6605
MS data file   : D:\Data\Lakshmy\120808\241108\SSP 6605_RC3_01_783.d\SSP 6605_RC3_01_783.mgf
Database       : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy       : Bacteria (Eubacteria) (5690016 sequences)
Timestamp      : 5 Jan 2010 at 12:01:23 GMT
Protein hits   : gi|261340146 2-dehydro-3-deoxyphosphooctonate aldolase [Enterobacter cancerogenus ATCC 35316]
                 gi|50582463  oxidoreductase [Exiguobacterium acetylicum]
                 gi|146311979 2-dehydro-3-deoxyphosphooctonate aldolase [Enterobacter sp. 638]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|261340146 **Mass:** 30917 **Score:** 144 **Queries matched:** 3 **emPAI:** 0.23
2-dehydro-3-deoxyphosphooctonate aldolase [Enterobacter cancerogenus ATCC 35316]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 194	723.3040	2166.8902	2167.0732	-0.1830	0	48	0.32	1	R.AGMATGLAGLFIEAHPDPANAK.C + Oxidation (M)
☑ 248	964.5330	2890.5772	2890.4899	0.0873	0	59	0.023	1	K.VVSIGDINVANDLPFVLFGGMNVLESR.D + Oxidation (M)
☑ 174	1034.5570	3100.6492	3100.6557	-0.0065	0	39	5.1	1	K.VITDVHEASQAQPVADVVDVIQLPAFLAR.Q

2. gi|50582463 **Mass:** 27360 **Score:** 118 **Queries matched:** 2 **emPAI:** 0.12
oxidoreductase [Exiguobacterium acetylicum]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 133	859.3740	1716.7334	1716.8570	-0.1236	0	86	6.4e-05	1	K.YTVITGASSGIGYETAK.L
☑ 89	677.2430	2028.7072	2028.9752	-0.2681	0	32	25	1	K.SVDLADNQNVHDLYEGLK.E

3. gi|146311979 **Mass:** 31019 **Score:** 107 **Queries matched:** 2 **emPAI:** 0.23
2-dehydro-3-deoxyphosphooctonate aldolase [Enterobacter sp. 638]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
194	723.3040	2166.8902	2167.0732	-0.1830	0	48	0.32	1	R.AGMATGIAGLFIEAHPDPANAK.C + Oxidation (M)
248	964.5330	2890.5772	2890.4899	0.0873	0	59	0.023	1	K.VVSIGDINVANDLPFVLFGGMNVLESR.D + Oxidation (M)

Peptide matches not assigned to protein hits: (no details means no match)

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 169	676.8940	2027.6602	2027.0510	0.6092	0	38	2.8	1	IALDQGMDIHSVFELLAR
☑ 85	651.8640	1301.7134	1301.6728	0.0407	1	34	14	1	VTDLDKAGAWAR
☑ 221	761.6660	2281.9762	2281.1028	0.8733	2	33	9.5	1	TNFLAGYRLEHRFNDDWK
☑ 138	592.3340	1773.9802	1773.9036	0.0765	0	33	14	1	VSAEEPLIADLFEVDK
☑ 128	846.1080	2535.3022	2535.1866	0.1156	1	30	37	1	IEAFFHGAGYAMHRHDTYAIGR + Oxidation (M)
☑ 246	953.0700	2856.1882	2856.2075	-0.0193	2	30	15	1	VAKCDQCDGDPACVKVCPTGALEFR
☑ 44	436.2640	870.5134	871.4763	-0.9628	0	29	42	1	EAALAELR
☑ 202	740.9910	2219.9512	2220.1678	-0.2166	0	29	28	1	IIDLDADYALTLVEPGVTFR
☑ 206	742.2930	2223.8572	2223.0049	0.8523	0	29	24	1	QMPAGYETQIGEGGGMLSGGQR
☑ 161	665.2980	1992.8722	1993.9745	-1.1023	0	28	33	1	ELIGAEDQFFEIGGNSLR

Mascot Search Results

Protein View

Match to: **gi|261340146** Score: **144**
2-dehydro-3-deoxyphosphooctonate aldolase [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\241108\SSP 6605_RC3_01_783.d\SSP 6605_RC3_01_783.mgf

Nominal mass (M_r): **30917**; Calculated pI value: **5.63**
NCBI BLAST search of gi|261340146 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **27%**

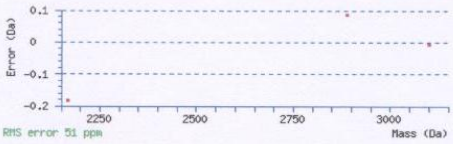
Matched peptides shown in **Bold Red**

```
  1 MKQKVVSIGD INVANDLPFV LFGGMNVLES RDLAMRICEH YVTVTQKLGI
 51 PYVFKASFDK ANRSSINSYR GPGLEEGMKI FQELKQTFGV KVITDVHEAS
101 QAQFVADVVD VIQLPAFLAR QTDLVAAMAK TGAVINVKKP QFVSPGQMGN
151 IVDKFIEGGN DKVILCDRGA NFGYDNLVVD MLGFSVMKNV SNQSPVIFDV
201 THALQCRDPF GAASGGRRAQ VTELARAGMA TGLAGLFIEA HPDPANAKCD
251 GPSALPLDKL EPFLKQIKAI DDLVKSFDEL DTSN
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
5 - 31	964.5330	2890.5772	2890.4899	0.0873	0	K.VVSIGDINVANDLPFVLFGGMNVLESR.D Oxidation (M) (Ions score 59)
92 - 120	1034.5570	3100.6492	3100.6557	-0.0065	0	K.VITDVHEASQAQFVADVVDVIQLPAFLAR.Q (Ions score 39)
227 - 248	723.3040	2166.8902	2167.0732	-0.1830	0	R.AGMATGLAGLFIEAHPDPANAK.C Oxidation (M) (Ions score 48)




```
LOCUS       ZP_05968004              284 aa            linear   BCT 15-OCT-2009
DEFINITION  2-dehydro-3-deoxyphosphooctonate aldolase [Enterobacter
            cancerogenus ATCC 35316].
ACCESSION   ZP_05968004
VERSION     ZP_05968004.1  GI:261340146
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000008.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 284)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 284)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000008.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
```

```
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..284
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..284
                     /product="2-dehydro-3-deoxyphosphooctonate aldolase"
                     /EC_number="2.5.1.55"
                     /calculated_mol_wt=30573
     Region          9..274
                     /region_name="NeuB"
                     /note="NeuB family; cl00496"
                     /db_xref="CDD:140752"
     Region          10..276
                     /region_name="DAHP_synth_1"
                     /note="DAHP synthetase I family; pfam00793"
                     /db_xref="CDD:109833"
     CDS             1..284
                     /locus_tag="EcanA3_020100006759"
                     /coded_by="complement(NZ_ABWM02000008.1:68..922)"
                     /note="COG2877 3-deoxy-D-manno-octulosonic acid (KDO)
                     8-phosphate synthase"
                     /transl_table=11
                     /db_xref="CDD:135264"
```

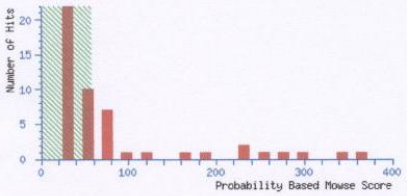
Mascot: http://www.matrixscience.com/

Mascot Search Results

User : Lakshmy Manickan
Email : lakshmy.manickan@unn.ac.uk
Search title : SSP 6711
MS data file : D:\Data\Lakshmy\120808\SSP 6711_RB5_01_717.d\SSP 6711_RB5_01_717.mgf
Database : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy : Bacteria (Eubacteria) (5690016 sequences)
Timestamp : 5 Jan 2010 at 12:02:31 GMT
Protein hits : gi|157147228 elongation factor Tu [Citrobacter koseri ATCC BAA-895]
gi|261819559 translation elongation factor Tu [Pectobacterium wasabiae WPP163]
gi|110807172 elongation factor Tu [Shigella flexneri 5 str. 8401]
gi|16272522 elongation factor Tu [Haemophilus influenzae Rd KW20]
gi|261341842 elongation factor Tu [Enterobacter cancerogenus ATCC 35316]
gi|259906922 Elongation factor Tu-A [Erwinia pyrifoliae Ep1/96]
gi|37524441 elongation factor Tu [Photorhabdus luminescens subsp. laumondii TTO1]
gi|237809513 translation elongation factor Tu [Tolumonas auensis DSM 9187]
gi|270341211 elongation factor Tu [Pectobacterium carotovorum subsp. carotovorum]
gi|78358029 elongation factor Tu [Desulfovibrio desulfuricans subsp. desulfuricans str. G20]
gi|24371815 elongation factor Tu [Shewanella oneidensis MR-1]
gi|78189809 elongation factor Tu [Chlorobium chlorochromatii CaD3]
gi|119468170 protein chain elongation factor EF-Tu; GTP-binding factor [Alteromonadales bacterium TW-7]
gi|2369692 elongation factor Ef-Tu [Buchnera aphidicola]
gi|88861465 protein chain elongation factor EF-Tu; GTP-binding factor [Pseudoalteromonas tunicata D2]
gi|19551739 elongation factor Tu [Corynebacterium glutamicum ATCC 13032]
gi|76786395 elongation factor Tu [endosymbiont of Haematomyzus elephantis]
gi|270341139 elongation factor Tu [Aeromonas molluscorum]
gi|57238897 elongation factor Tu [Ehrlichia ruminantium str. Welgevonden]
gi|15839217 elongation factor Tu [Xylella fastidiosa 9a5c]
gi|119947040 elongation factor Tu [Psychromonas ingrahamii 37]

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As Peptide Summary Help
Significance threshold p< 0.05 Max. number of hits AUTO
Standard scoring MudPIT scoring Ions score or expect cut-off 0 Show sub-sets 0
Show pop-ups Suppress pop-ups Sort unassigned Decreasing Score Require bold red
Select All Select None Search Selected Error tolerant

1. gi|157147228 **Mass:** 43389 **Score:** 365 **Queries matched:** 10 **emPAI:** 0.67
elongation factor Tu [Citrobacter koseri ATCC BAA-895]
Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
78	652.2320	1302.4494	1302.7759	-0.3264	0	62	0.014	1	K.TTLTAAITTVLAK.T
136	604.9370	1811.7892	1811.9274	-0.1382	0	39	2.8	1	K.MVVTLIHPIAMDDGLR.F + 2 Oxidation (M)
164	982.9230	1963.8314	1963.9527	-0.1213	0	61	0.019	1	R.ELLSQYDFPGDDTPIVR.G
166	989.4440	1976.8734	1977.0095	-0.1360	0	40	2.6	1	K.IIELAGYLDSYIPEPER.A
167	989.7760	1977.5374	1977.0095	0.5280	0	(25)	99	1	K.IIELAGYLDSYIPEPER.A
194	706.3230	2115.9472	2116.1568	-0.2097	0	(33)	11	1	R.AIDKPFLLPIEDVFSISGR.G
195	706.3510	2116.0312	2116.1568	-0.1257	0	(48)	0.42	1	R.AIDKPFLLPIEDVFSISGR.G
196	1059.0410	2116.0674	2116.1568	-0.0894	0	52	0.15	1	R.AIDKPFLLPIEDVFSISGR.G
241	860.0200	2577.0382	2577.2190	-0.1808	0	54	0.066	1	R.TTDVTGTIELPEGVEMVMPGDNIK.M + 2 Oxidation (M)
246	926.6500	2776.9282	2776.3194	0.6088	0	57	0.026	1	K.NMITGAAQMDGAILVVAATDGPMPQTR.E + 3 Oxidation (M)

Proteins matching the same set of peptides:
gi|157148907 **Mass:** 45003 **Score:** 365 **Queries matched:** 10
elongation factor Tu [Citrobacter koseri ATCC BAA-895]

2. gi|261819559 **Mass:** 43374 **Score:** 338 **Queries matched:** 9 **emPAI:** 0.55
translation elongation factor Tu [Pectobacterium wasabiae WPP163]
Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
78	652.2320	1302.4494	1302.7759	-0.3264	0	62	0.014	1	K.TTLTAAITTVLAK.T
99	**737.4140**	**1472.8134**	**1472.8755**	**-0.0621**	**0**	**14**	**1.2e+03**	**6**	**R.QVGVPFIIVFLNK.C**
136	604.9370	1811.7892	1811.9274	-0.1382	0	39	2.8	1	K.MVVTLIHPIAMDDGLR.F + 2 Oxidation (M)
164	982.9230	1963.8314	1963.9527	-0.1213	0	61	0.019	1	R.ELLSQYDFPGDDTPVIR.G
194	706.3230	2115.9472	2116.1568	-0.2097	0	(33)	11	1	R.AIDKPFLLPIEDVFSISGR.G
195	706.3510	2116.0312	2116.1568	-0.1257	0	(48)	0.42	1	R.AIDKPFLLPIEDVFSISGR.G
196	1059.0410	2116.0674	2116.1568	-0.0894	0	52	0.15	1	R.AIDKPFLLPIEDVFSISGR.G
241	860.0200	2577.0382	2577.2190	-0.1808	0	54	0.066	1	R.TTDVTGTIELPEGVEMVMPGDNIK.M + 2 Oxidation (M)
246	926.6500	2776.9282	2776.3194	0.6088	0	57	0.026	1	K.NMITGAAQMDGAILVVAATDGPMPQTR.E + 3 Oxidation (M)

3. gi|110807172 **Mass:** 43429 **Score:** 304 **Queries matched:** 8 **emPAI:** 0.44
elongation factor Tu [Shigella flexneri 5 str. 8401]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
78	652.2320	1302.4494	1302.7759	-0.3264	0	62	0.014	1	K.TTLTAAITTVLAK.T
136	604.9370	1811.7892	1811.9274	-0.1382	0	39	2.8	1	K.MVVTLIHPIAMDDGLR.F + 2 Oxidation (M)
164	982.9230	1963.8314	1963.9527	-0.1213	0	61	0.019	1	R.ELLSQYDFPGDDTPIVR.G
194	706.3230	2115.9472	2116.1568	-0.2097	0	(33)	11	1	R.AIDKPFLLPIEDVFSISGR.G
195	706.3510	2116.0312	2116.1568	-0.1257	0	(48)	0.42	1	R.AIDKPFLLPIEDVFSISGR.G
196	1059.0410	2116.0674	2116.1568	-0.0894	0	52	0.15	1	R.AIDKPFLLPIEDVFSISGR.G
241	860.0200	2577.0382	2576.2350	0.8032	0	32	11	2	R.TTDVTGTIELPEGVQMVMPGDNIK.M + 2 Oxidation (M)
246	926.6500	2776.9282	2776.3194	0.6088	0	57	0.026	1	K.NMITGAAQMDGAILVVAATDGPMPQTR.E + 3 Oxidation (M)

4. gi|16272522 **Mass:** 43443 **Score:** 286 **Queries matched:** 7 **emPAI:** 0.55
elongation factor Tu [Haemophilus influenzae Rd KW20]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
78	652.2320	1302.4494	1302.7759	-0.3264	0	62	0.014	1	K.TTLTAAITTVLAK.H
164	982.9230	1963.8314	1963.9527	-0.1213	0	61	0.019	1	R.ELLSQYDFPGDDTPIVR.G
194	706.3230	2115.9472	2116.1205	-0.1733	0	(33)	11	1	R.AIDQPFLLPIEDVFSISGR.G
195	706.3510	2116.0312	2116.1205	-0.0893	0	(48)	0.42	1	R.AIDQPFLLPIEDVFSISGR.G
196	1059.0410	2116.0674	2116.1205	-0.0530	0	52	0.15	1	R.AIDQPFLLPIEDVFSISGR.G
241	860.0200	2577.0382	2577.2190	-0.1808	0	54	0.066	1	R.TTDVTGTIELPEGVEMVMPGDNIK.M + 2 Oxidation (M)
246	926.6500	2776.9282	2776.3194	0.6088	0	57	0.026	1	K.NMITGAAQMDGAILVVAATDGPMPQTR.E + 3 Oxidation (M)

Proteins matching the same set of peptides:
gi|53733337 **Mass:** 43469 **Score:** 286 **Queries matched:** 7
COG0050: GTPases - translation elongation factors [Haemophilus influenzae R2866]
gi|145629756 **Mass:** 43442 **Score:** 286 **Queries matched:** 7
tRNA-dihydrouridine synthase A [Haemophilus influenzae 22.1-21]
gi|145640555 **Mass:** 43471 **Score:** 286 **Queries matched:** 7
hypothetical protein CGSHiR3021_10860 [Haemophilus influenzae R3021]
gi|15603222 **Mass:** 43442 **Score:** 284 **Queries matched:** 7
elongation factor Tu [Pasteurella multocida subsp. multocida str. Pm70]
gi|42630895 **Mass:** 43386 **Score:** 284 **Queries matched:** 7
COG0050: GTPases - translation elongation factors [Haemophilus influenzae R2866]
gi|251793286 **Mass:** 43438 **Score:** 284 **Queries matched:** 7
translation elongation factor Tu [Aggregatibacter aphrophilus NJ8700]

5. gi|261341842 **Mass:** 38502 **Score:** 246 **Queries matched:** 7 **emPAI:** 0.51
elongation factor Tu [Enterobacter cancerogenus ATCC 35316]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
78	652.2320	1302.4494	1302.7759	-0.3264	0	62	0.014	1	K.TTLTAAITTVLAK.T
99	737.4140	1472.8134	1472.8755	-0.0621	0	14	1.2e+03	6	R.QVGVPFIIVFLNK.C
164	982.9230	1963.8314	1963.9527	-0.1213	0	61	0.019	1	R.ELLSQYDFPGDDTPIVR.G
194	706.3230	2115.9472	2116.1568	-0.2097	0	(33)	11	1	R.AIDKPFLLPIEDVFSISGR.G
195	706.3510	2116.0312	2116.1568	-0.1257	0	(48)	0.42	1	R.AIDKPFLLPIEDVFSISGR.G
196	1059.0410	2116.0674	2116.1568	-0.0894	0	52	0.15	1	R.AIDKPFLLPIEDVFSISGR.G
246	926.6500	2776.9282	2776.3194	0.6088	0	57	0.026	1	K.NMITGAAQMDGAILVVAATDGPMPQTR.E + 3 Oxidation (M)

6. gi|259906922 **Mass:** 43465 **Score:** 233 **Queries matched:** 7 **emPAI:** 0.34
Elongation factor Tu-A [Erwinia pyrifoliae Ep1/96]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
78	652.2320	1302.4494	1302.7759	-0.3264	0	62	0.014	1	K.TTLTAAITTVLAK.T
99	737.4140	1472.8134	1472.8755	-0.0621	0	14	1.2e+03	6	R.QVGVPFIIVFLNK.C
164	982.9230	1963.8314	1963.9527	-0.1213	0	47	0.45	4	R.DLLTQYDFPGDDTPIVR.G
194	706.3230	2115.9472	2116.1568	-0.2097	0	(33)	11	1	R.AIDKPFLLPIEDVFSISGR.G
195	706.3510	2116.0312	2116.1568	-0.1257	0	(48)	0.42	1	R.AIDKPFLLPIEDVFSISGR.G
196	1059.0410	2116.0674	2116.1568	-0.0894	0	52	0.15	1	R.AIDKPFLLPIEDVFSISGR.G
246	926.6500	2776.9282	2776.3194	0.6088	0	57	0.026	1	K.NMITGAAQMDGAILVVAATDGPMPQTR.E + 3 Oxidation (M)

7. gi|37524441 **Mass:** 43321 **Score:** 232 **Queries matched:** 6 **emPAI:** 0.44
elongation factor Tu [Photorhabdus luminescens subsp. laumondii TT01]
☐ Check to include this hit in error tolerant search

Mascot Search Results

Protein View

Match to: **gi|261341842** Score: **246**
elongation factor Tu [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\SSP 6711_RB5_01_717.d\SSP 6711_RB5_01_717.mgf

Nominal mass (M_r): **38502**; Calculated pI value: **5.17**
NCBI BLAST search of gi|261341842 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **25%**

Matched peptides shown in **Bold Red**

```
  1 MSKEKFERTK PHVNVGTIGH VDHGKTTLTA AITTVLAKTY GGAARAFDQI
 51 DNAPEEKARG ITINTSHVEY DTPTRHYAHV DCPGHADYVK NMITGAAQMD
101 GAILVVAATD GPMPQTREHI LLGRQVGVPF IIVFLNKCDM VDDEELLELV
151 EMEVRELLSQ YDFPGDDTPI VRGSALKALE GEAEWEEKII ELAGFLDSYI
201 PEPERAIDKP FLLPIEDVFS ISGRGTVVTG RVERGIIKVG EEVEIVGIKE
251 TAKSTCTGVE MFRKLLDEGR AGENVGVLLR GIKREEIERG QVLAKPGSIK
301 PHTKFESEVY ILSKDEGGRH TPFFKGYRPQ FYFRTTDVTG TIELPEG
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
26 - 38	652.2320	1302.4494	1302.7759	-0.3264	0	K.TTLTAAITTVLAK.T (Ions score 62)
91 - 117	926.6500	2776.9282	2776.3194	0.6088	0	K.NMITGAAQMDGAILVVAATDGPMPQTR.E 3 Oxidation (M) (Ions score 57)
125 - 137	737.4140	1472.8134	1472.8755	-0.0621	0	R.QVGVPFIIVFLNK.C (Ions score 14)
156 - 172	982.9230	1963.8314	1963.9527	-0.1213	0	R.ELLSQYDFPGDDTPIVR.G (Ions score 61)
206 - 224	706.3230	2115.9472	2116.1568	-0.2097	0	R.AIDKPFLLPIEDVFSISGR.G (Ions score 33)
206 - 224	706.3510	2116.0312	2116.1568	-0.1257	0	R.AIDKPFLLPIEDVFSISGR.G (Ions score 48)
206 - 224	1059.0410	2116.0674	2116.1568	-0.0894	0	R.AIDKPFLLPIEDVFSISGR.G (Ions score 52)

Error: try setting browser cache to automatic. Error: try setting browser cache to automatic.

```
LOCUS       ZP_05969700              347 aa            linear   BCT 15-OCT-2009
DEFINITION  elongation factor Tu [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05969700
VERSION     ZP_05969700.1  GI:261341842
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000029.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 347)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 347)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000029.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
```

```
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            COMPLETENESS: incomplete on the carboxy end.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..347
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..>347
                     /product="elongation factor Tu"
                     /EC_number="3.6.5.3"
                     /calculated_mol_wt=38221
     Region          11..203
                     /region_name="EF_Tu"
                     /note="EF-Tu subfamily.  This subfamily includes orthologs
                     of translation elongation factor EF-Tu in bacteria,
                     mitochondria, and chloroplasts.  It is one of several
                     GTP-binding translation factors found in the larger family
                     of GTP-binding elongation factors...; cd01884"
                     /db_xref="CDD:133284"
     Site            19..26
                     /site_type="other"
                     /note="G1 box"
                     /db_xref="CDD:133284"
     Site            order(20,22,26..27,30,33..34,38,66..67,85..86,109..110,
                     112..115,118,149,153,177..178)
                     /site_type="other"
                     /note="GEF interaction site"
                     /db_xref="CDD:133284"
     Site            order(22..27,47,136..137,139,174)
                     /site_type="other"
                     /note="GTP/Mg2+ binding site"
                     /db_xref="CDD:133284"
     Site            56..66
                     /site_type="other"
                     /note="Switch I region"
                     /db_xref="CDD:133284"
     Site            62
                     /site_type="other"
                     /note="G2 box"
                     /db_xref="CDD:133284"
     Site            81..84
                     /site_type="other"
                     /note="G3 box"
                     /db_xref="CDD:133284"
     Site            83..101
                     /site_type="other"
                     /note="Switch II region"
                     /db_xref="CDD:133284"
     Site            136..139
                     /site_type="other"
                     /note="G4 box"
                     /db_xref="CDD:133284"
     Region          211..297
                     /region_name="EFTU_II"
                     /note="EFTU_II: Elongation factor Tu domain II. Elongation
                     factors Tu (EF-Tu) are three-domain GTPases with an
                     essential function in the elongation phase of mRNA
                     translation. The GTPase center of EF-Tu is in the
                     N-terminal domain (domain I), also known as the...;
                     cd03697"
                     /db_xref="CDD:58088"
     Region          300..>347
                     /region_name="Translation_factor_III"
                     /note="Domain III of Elongation factor (EF) Tu (EF-TU) and
                     EF-G.  Elongation factors (EF) EF-Tu and EF-G participate
                     in the elongation phase during protein biosynthesis on the
                     ribosome. Their functional cycles depend on GTP binding
                     and its hydrolysis. The EF...; cl02786"
                     /db_xref="CDD:121545"
     CDS             1..347
                     /locus_tag="EcanA3_020100015362"
                     /coded_by="complement(NZ_ABWM02000029.1:<1..1040)"
                     /note="COG0050 GTPases - translation elongation factors"
                     /transl_table=11
                     /db_xref="CDD:134063"
```

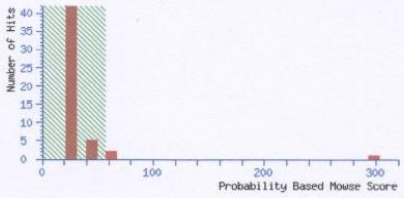
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 7602
MS data file    : D:\Data\Lakshmy\120808\260808\SSP 7602_RH1_01_413.d\SSP 7602_RH1_01_413.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 5 Jan 2010 at 12:05:52 GMT
Protein hits    : gi|146311681 acetoin reductase [Enterobacter sp. 638]
                  gi|2285882   Aalpha subunit of acetyl-CoA carboxylase [Escherichia coli W3110]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|146311681 **Mass:** 26942 **Score:** 298 **Queries matched:** 5 **emPAI:** 0.42
acetoin reductase [Enterobacter sp. 638]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	104	759.0050	1515.9954	1515.8198	0.1757	0	97	4.7e-06	1	K.GVIWGIQAAIDAFR.K
☑	128	564.6070	1690.7992	1689.8686	0.9306	1	41	2.1	1	K.VDVSNREQVFAAVEK.A
☑	132	571.8780	1712.6122	1712.7893	-0.1772	0	(53)	0.089	1	K.DGFAVAIADYNEETAK.A
☑	133	857.4090	1712.8034	1712.7893	0.0141	0	88	4e-05	1	K.DGFAVAIADYNEETAK.A
☑	234	802.0650	2403.1732	2404.1329	-0.9597	0	72	0.0014	1	K.IINACSQAGHTGNPELAVYSSSK.F

Proteins matching the same set of peptides:
gi|261341164 **Mass:** 26884 **Score:** 298 **Queries matched:** 5
acetoin reductase [Enterobacter cancerogenus ATCC 35316]

---

2. gi|2285882 **Score:** 59 **Queries matched:** 2
Aalpha subunit of acetyl-CoA carboxylase [Escherichia coli W3110]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
111	529.6000	1585.7782	1586.8780	-1.0998	0	32	18	3	K.LIDSIIPEPLGGAHR.N
143	599.9910	1796.9512	1796.9454	0.0058	1	27	59	3	K.SADKAPLAAEAMGIIAPR.L + Oxidation (M)

Proteins matching the same set of peptides:
gi|15799867 **Score:** 59 **Queries matched:** 2
gi|16759222 **Score:** 59 **Queries matched:** 2
gi|26246132 **Score:** 59 **Queries matched:** 2
gi|74310805 **Score:** 59 **Queries matched:** 2
gi|82542784 **Score:** 59 **Queries matched:** 2
gi|91209255 **Score:** 59 **Queries matched:** 2
gi|161504649 **Score:** 59 **Queries matched:** 2
gi|170768662 **Score:** 59 **Queries matched:** 2
gi|187731373 **Score:** 59 **Queries matched:** 2
gi|213029802 **Score:** 59 **Queries matched:** 2
gi|213416583 **Score:** 59 **Queries matched:** 2
gi|213586740 **Score:** 59 **Queries matched:** 2
gi|224582080 **Score:** 59 **Queries matched:** 2

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	111	529.6000	1585.7782	1586.8780	-1.0999	0	40	3	1	LIDTVIPEPLGGAHR

Mascot Search Results

Protein View

Match to: **gi|261341164** Score: **298**
acetoin reductase [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\260808\SSP 7602_RH1_01_413.d\SSP 7602_RH1_01_413.mgf

Nominal mass (M_r): **26884**; Calculated pI value: **6.17**
NCBI BLAST search of gi|261341164 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **26%**

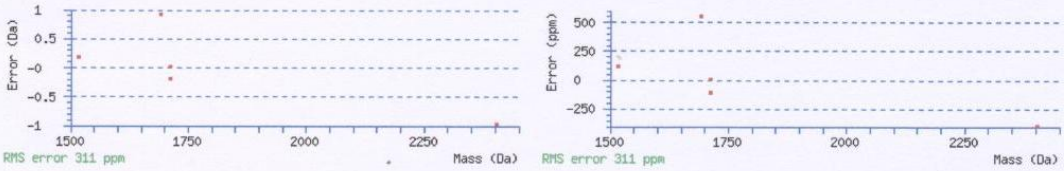
Matched peptides shown in **Bold Red**

```
  1 MQKVALVTGS GQGIGKAIAL RLVKDGFAVA IADYNEETAK AVAEEIVRSG
 51 GKAVAVKVDV SNREQVFAAV EKARTALGGF NVIVNNAGVA PSTPIESITP
101 DIVDKVYNIN VKGVIWGIQA AIDAFRKEGH GGKIINACSQ AGHTGNPELA
151 VYSSSKFAVR GLTQTAARDL APLGITVNAY CPGIVKTPMW AEIDRQVSEA
201 AGKPLGYGTE TFAKRITLGR LSEPEDVAAC VSYLAGPDSD YMTGQSLLID
251 GGMVFN
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
25 - 40	571.8780	1712.6122	1712.7893	-0.1772	0	K.DGFAVAIADYNEETAK.A (Ions score 53)
25 - 40	857.4090	1712.8034	1712.7893	0.0141	0	K.DGFAVAIADYNEETAK.A (Ions score 88)
58 - 72	564.6070	1690.7992	1689.8686	0.9306	1	K.VDVSNREQVFAAVEK.A (Ions score 41)
113 - 126	759.0050	1515.9954	1515.8198	0.1757	0	K.GVIWGIQAAIDAFR.K (Ions score 97)
134 - 156	802.0650	2403.1732	2404.1329	-0.9597	0	K.IINACSQAGHTGNPELAVYSSSK.F (Ions score 72)



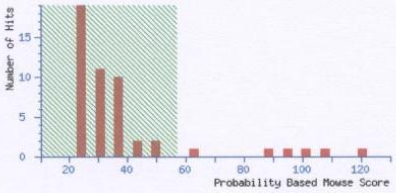

```
LOCUS       ZP_05969022              256 aa            linear   BCT 15-OCT-2009
DEFINITION  acetoin reductase [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05969022
VERSION     ZP_05969022.1  GI:261341164
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000020.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 256)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 256)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000020.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
```

Mascot Search Results

```
User          : Lakshmy Manickan
Email         : lakshmy.manickan@unn.ac.uk
Search title  : SSP 7707
MS data file  : D:\Data\Lakshmy\120808\280808\SSP 7707_RI2_01_457.d\SSP 7707_RI2_01_457.mgf
Database      : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy      : Bacteria (Eubacteria) (5690016 sequences)
Timestamp     : 5 Jan 2010 at 12:11:18 GMT
Protein hits  : gi|157368709 inorganic pyrophosphatase [Serratia proteamaculans 568]
                gi|269137738 Inorganic pyrophosphatase [Edwardsiella tarda EIB202]
                gi|6453807   ProA [Shigella flexneri]
                gi|1943414   Chain A, Inorganic Pyrophosphatase
                gi|258634603 gamma-glutamyl phosphate reductase [Pantoea sp. At-9b]
                gi|2781250   Chain A, Structure Of Inorganic Pyrophosphatase Mutant D42n
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned Decreasing Score Require bold red ☐

Select All | Select None | Search Selected ☐ Error tolerant

1. gi|157368709 **Mass:** 19679 **Score:** 120 **Queries matched:** 4
inorganic pyrophosphatase [Serratia proteamaculans 568]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	68	518.2850	1034.5554	1034.5396	0.0158	0	47	0.62	1	K.AEIIASFER.A
	73	578.3370	1154.6594	1154.6183	0.0412	0	33	13	3	K.DVNDLPELLK.A
☑	130	547.9490	1640.8252	1640.8046	0.0206	1	40	2.8	1	K.YEIDKDTGALFVDR.F
☑	131	821.4700	1640.9254	1640.8046	0.1209	1	(15)	8.4e+02	1	K.YEIDKDTGALFVDR.F

2. gi|269137738 **Mass:** 19766 **Score:** 107 **Queries matched:** 4
Inorganic pyrophosphatase [Edwardsiella tarda EIB202]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
68	518.2850	1034.5554	1034.5396	0.0158	0	47	0.62	1	K.AEIIASFER.A
85	464.2630	1389.7672	1389.6830	0.0842	0	20	3.1e+02	3	R.AQIAHFFEHYK.D
130	547.9490	1640.8252	1640.8046	0.0206	1	40	2.8	1	K.YEIDKDTGALFVDR.F
131	821.4700	1640.9254	1640.8046	0.1209	1	(15)	8.4e+02	1	K.YEIDKDTGALFVDR.F

3. gi|6453807 **Mass:** 22056 **Score:** 99 **Queries matched:** 2 **emPAI:** 0.15
ProA [Shigella flexneri]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	77	648.8340	1295.6534	1295.6431	0.0103	0	70	0.003	1	R.GPMGLEALTTYK.W + Oxidation (M)
☑	183	619.9010	1856.6812	1856.8905	-0.2093	0	29	24	1	R.FVNEVDSSAVYVNASTR.F

Proteins matching the same set of peptides:
gi|15799948 **Mass:** 45102 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Escherichia coli O157:H7 EDL933]
gi|16128229 **Mass:** 45008 **Score:** 99 **Queries matched:** 2
gamma-glutamylphosphate reductase [Escherichia coli str. K-12 substr. MG1655]
gi|24111735 **Mass:** 45010 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Shigella flexneri 2a str. 301]
gi|26246288 **Mass:** 45012 **Score:** 99 **Queries matched:** 2

```
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..256
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..256
                     /product="acetoin reductase"
                     /calculated_mol_wt=26595
     Region          1..256
                     /region_name="PRK08643"
                     /note="acetoin reductase; Validated; PRK08643"
                     /db_xref="CDD:136931"
     Region          1..253
                     /region_name="NADB_Rossmann"
                     /note="Rossmann-fold NAD(P)(+)-binding proteins; cl09931"
                     /db_xref="CDD:143725"
     CDS             1..256
                     /locus_tag="EcanA3_020100011911"
                     /coded_by="NZ_ABWM02000020.1:70258..71028"
                     /note="COG1028 Dehydrogenases with different specificities
                     (related to short-chain alcohol dehydrogenases)"
                     /transl_table=11
                     /db_xref="CDD:136931"
```

Mascot: http://www.matrixscience.com/

gamma-glutamyl phosphate reductase [Escherichia coli CFT073]
gi|82542846 **Mass:** 45013 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Shigella boydii Sb227]
gi|82775822 **Mass:** 44980 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Shigella dysenteriae Sd197]
gi|91209332 **Mass:** 44954 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Escherichia coli UTI89]
gi|110640479 **Mass:** 44959 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Escherichia coli 536]
gi|157147174 **Mass:** 44879 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Citrobacter koseri ATCC BAA-895]
gi|157157629 **Mass:** 45008 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Escherichia coli E24377A]
gi|170682591 **Mass:** 45038 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Escherichia coli SMS-3-5]
gi|170767876 **Mass:** 45086 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Escherichia albertii TW07627]
gi|188534718 **Mass:** 45397 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Erwinia tasmaniensis Et1/99]
gi|194432663 **Mass:** 44978 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Shigella dysenteriae 1012]
gi|194438160 **Mass:** 45038 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Escherichia coli 101-1]
gi|215485391 **Mass:** 45012 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Escherichia coli O127:H6 str. E2348/69]
gi|218550050 **Mass:** 44924 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Escherichia fergusonii ATCC 35469]
gi|218688135 **Mass:** 45012 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Escherichia coli ED1a]
gi|218693726 **Mass:** 45039 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Escherichia coli 55989]
gi|218698813 **Mass:** 45038 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Escherichia coli IAI39]
gi|218703559 **Mass:** 44996 **Score:** 99 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Escherichia coli UMN026]
gi|237730224 **Mass:** 45044 **Score:** 99 **Queries matched:** 2
gamma-glutamylphosphate reductase [Citrobacter sp. 30_2]
gi|260866441 **Mass:** 44994 **Score:** 99 **Queries matched:** 2
gamma-glutamylphosphate reductase [Escherichia coli O111:H- str. 11128]
gi|281177456 **Mass:** 45000 **Score:** 99 **Queries matched:** 2
gamma-glutamylphosphate reductase [Escherichia coli SE15]
gi|156935257 **Mass:** 45461 **Score:** 97 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Enterobacter sakazakii ATCC BAA-894]
gi|259909344 **Mass:** 45425 **Score:** 97 **Queries matched:** 2
Gamma-glutamyl phosphate reductase [Erwinia pyrifoliae Ep1/96]
gi|260596647 **Mass:** 45456 **Score:** 97 **Queries matched:** 2
Gamma-glutamyl phosphate reductase [Cronobacter turicensis]
gi|261340945 **Mass:** 45230 **Score:** 97 **Queries matched:** 2
gamma-glutamyl phosphate reductase [Enterobacter cancerogenus ATCC 35316]

---

4. gi|1943414 **Score:** 93 **Queries matched:** 3
Chain A, Inorganic Pyrophosphatase
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
73	578.3370	1154.6594	1154.6183	0.0412	0	33	13	3	K.DVNDLPELLK.A
85	464.2630	1389.7672	1389.6830	0.0842	0	20	3.1e+02	3	K.AQIAHFFEHYK.D
130	547.9490	1640.8252	1640.8046	0.0206	1	40	2.8	1	K.YEIDKESGALFVDR.F

Proteins matching the same set of peptides:
gi|2554846 **Score:** 93 **Queries matched:** 3
gi|2781252 **Score:** 93 **Queries matched:** 3
gi|2781254 **Score:** 93 **Queries matched:** 3
gi|15804817 **Score:** 93 **Queries matched:** 3
gi|26251130 **Score:** 93 **Queries matched:** 3
gi|30065484 **Score:** 93 **Queries matched:** 3
gi|157831975 **Score:** 93 **Queries matched:** 3
gi|170766732 **Score:** 93 **Queries matched:** 3
gi|215489572 **Score:** 93 **Queries matched:** 3
gi|218551521 **Score:** 93 **Queries matched:** 3
gi|256019879 **Score:** 93 **Queries matched:** 3

---

5. gi|258634603 **Mass:** 45492 **Score:** 88 **Queries matched:** 2 **emPAI:** 0.07
gamma-glutamyl phosphate reductase [Pantoea sp. At-9b]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
77	648.8340	1295.6534	1295.6431	0.0103	0	70	0.003	1	R.GPMGLEALTTYK.W + Oxidation (M)
183	619.9010	1856.6812	1855.9064	0.7747	0	18	3.1e+02	3	R.FVNQVDSSAVYVNASTR.F

---

6. gi|2781250 **Score:** 61 **Queries matched:** 3
Chain A, Structure Of Inorganic Pyrophosphatase Mutant D42n
☐ Check to include this hit in error tolerant search

Mascot Search Results

Protein View

Match to: **gi|261340945** Score: **97**
gamma-glutamyl phosphate reductase [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\280808\SSP 7707_RI2_01_457.d\SSP 7707_RI2_01_457.mgf

Nominal mass (M_r): **45230**; Calculated pI value: **5.60**
NCBI BLAST search of gi|261340945 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **6%**

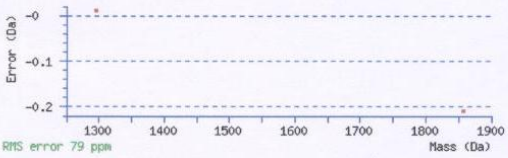
Matched peptides shown in **Bold Red**

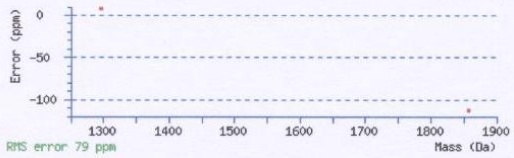
```
  1 MLEQMGAAAK AASYKLALLS SREKNRVLEK IADYLESQAS EILLANEQDL
 51 LEARRNGLSE AMLDRLALNP ARLKSIADDV RQVCNLADPV GQVIDGGLLD
101 SGLRLERRRV PLGVIGVIYE ARPNVTVDVA SLCLKTGNAA ILRGGKETWR
151 TNAATVKVIQ QALEECGLPA GAVQAIESPD RALVNEMLRM DKYIDMLIPR
201 GGAGLHKLCR EQSTIPVITG GIGVCHILVD DTAEIDPALK IIVNAKTQRP
251 STCNTVETLL VHQGIARTFL PALSQQMAQS GVTLHADASA LALLKDGPAN
301 VVPVKAEQYD DEFLSLDLNV KVVADLDDAI AHIREHGTQH SDAILTRTLR
351 NANRFVNEVD SSAVYVNAST RFTDGGQFGL GAEVAVSTQK LHARGPMGLE
401 ALTTYKWIGF GDDTIRA
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
355 - 371	619.9010	1856.6812	1856.8905	-0.2093	0	R.FVNEVDSSAVYVNASTR.F (Ions score 29)
395 - 406	648.8340	1295.6534	1295.6431	0.0103	0	R.GPMGLEALTTYK.W Oxidation (M) (Ions score 70)






```
LOCUS       ZP_05968803              417 aa            linear   BCT 15-OCT-2009
DEFINITION  gamma-glutamyl phosphate reductase [Enterobacter cancerogenus ATCC
            35316].
ACCESSION   ZP_05968803
VERSION     ZP_05968803.1  GI:261340945
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000017.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 417)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 417)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000017.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
```

```
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..417
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..417
                     /product="gamma-glutamyl phosphate reductase"
                     /EC_number="1.2.1.41"
                     /calculated_mol_wt=44778
     Region          22..412
                     /region_name="ALDH_F18-19_ProA-GPR"
                     /note="Gamma-glutamyl phosphate reductase (GPR), aldehyde
                     dehydrogenase families 18 and 19; cd07079"
                     /db_xref="CDD:143398"
     Region          22..406
                     /region_name="proA"
                     /note="gamma-glutamyl phosphate reductase; TIGR00407"
                     /db_xref="CDD:129501"
     Site            253
                     /site_type="other"
                     /note="putative catalytic cysteine"
                     /db_xref="CDD:143398"
     CDS             1..417
                     /gene="proA"
                     /locus_tag="EcanA3_020100010808"
                     /coded_by="NZ_ABWM02000017.1:31515..32768"
                     /note="COG0014 Gamma-glutamyl phosphate reductase"
                     /transl_table=11
                     /db_xref="CDD:134165"
```

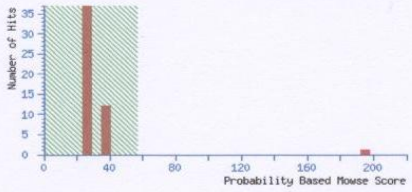
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 8205
MS data file    : D:\Data\Lakshmy\120808\280808\SSP 8205_RJ23_01_550.d\SSP 8205_RJ23_01_550.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 5 Jan 2010 at 12:12:48 GMT
Protein hits    : gi|261339072 YfaZ family protein [Enterobacter cancerogenus ATCC 35316]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|261339072 **Mass:** 18559 **Score:** 195 **Queries matched:** 3 **emPAI:** 0.65
   YfaZ family protein [Enterobacter cancerogenus ATCC 35316]
   ☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	95	648.8060	1295.5974	1295.6180	-0.0205	0	55	0.084	1	K.NFTNLNLEMGK.S + Oxidation (M)
☑	106	721.3580	1440.7014	1440.6885	0.0130	0	61	0.022	1	K.SSSGLYAESNWLK.N
☑	185	678.2780	2031.8122	2030.9810	0.8311	0	79	0.00025	1	K.DGRPNHTLIDGAYVGGGVSF.-

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	82	554.8700	1107.7254	1106.5720	1.1535	1	38	8.2	1	LAEAGEFSKR
☑	74	516.2960	1030.5774	1030.5546	0.0228	0	37	6.5	1	VLDELDSLK
☑	88	590.3090	1178.6034	1178.6546	-0.0512	2	33	14	1	YEELLKEKK
☑	216	737.6970	2210.0692	2210.0062	0.0630	2	32	14	1	NGDLMKEWTFDEVRANAAQ + Oxidation (M)
☑	225	747.0420	2238.1042	2237.0646	1.0396	1	32	15	1	GLGPAAGGRGGPGGGAGSGGTGGSGGAAGR
☑	229	762.0610	2283.1612	2283.0664	0.0948	0	31	20	1	MESPGAVSLWFGNGTSEMVLR + Oxidation (M)
☑	232	767.3850	2299.1332	2300.0896	-0.9564	1	30	23	1	FLRNYHTDFEVSNVCAISK
☑	81	553.5800	1105.1454	1105.5437	-0.3983	0	30	50	1	DLNTMLDLR + Oxidation (M)
☑	112	738.3990	1474.7834	1475.8347	-1.0513	2	29	36	1	FNKEQLDIIKTK
☑	92	621.9790	1241.9434	1242.7295	-0.7861	2	28	71	1	ILETNKELKR
☑	217	737.7010	2210.0812	2211.0598	-0.9787	0	28	39	1	EHPRPPVMMSTASQMLVER + Oxidation (M)
☑	201	1065.5450	2129.0754	2128.9657	0.1098	0	27	77	1	TFDNGMICASEQSVIVVDK + Oxidation (M)
☑	78	536.2960	1070.5774	1069.4862	1.0912	0	27	1.1e+02	1	MIDNYSTAR
☑	42	435.7910	869.5674	870.5902	-1.0228	1	27	55	1	KATVLVLK
☑	94	633.2510	1264.4874	1265.5267	-1.0393	0	27	89	1	QEMDMENEIK
☑	17	386.0810	770.1474	769.4195	0.7280	1	26	45	1	AGVDPRR
☑	139	876.3290	2625.9652	2625.2964	0.6687	2	26	84	1	LAKGMILDPSELISMSDFLRGCR + Oxidation (M)
☑	38	428.7700	855.5254	855.4814	0.0440	0	26	75	1	LATLHSSK
☑	220	742.3550	2224.0432	2223.0240	1.0192	2	26	60	1	MPRGGAFAWGGSRTQTDAEGR + Oxidation (M)
☑	37	428.6860	855.3574	855.5793	-0.2219	0	26	81	1	IATVLVIK
☑	156	934.9720	2801.8942	2802.2524	-0.3582	2	25	1.3e+02	1	RIHHIMVECSDPNDVGLAYDRCK + Oxidation (M)
☑	207	720.4810	2158.4212	2158.1780	0.2432	1	25	56	1	GINQLDLRQLMATSLSAVTK
☑	103	704.6680	2110.9822	2112.0851	-1.1030	1	25	1.4e+02	1	VTDLNGPGFDVPADALDRLK
☑	248	927.6190	2779.8352	2779.3466	0.4885	1	25	49	1	EIHNSNLDFPPSNHSFRVEHYLK
☑	238	788.8070	2363.3992	2364.1743	-0.7752	1	24	69	1	KQGPTSVAYVEVNNNSMLNVGK + Oxidation (M)
☑	233	1155.8640	3464.5702	3465.6441	-1.0739	1	24	1.1e+02	1	QTMTNMMKYLHQVMPPFIPADVIDDIETR + 2 Oxidation (M)
☑	69	496.0800	1485.2182	1485.8766	-0.6584	2	24	1.6e+02	1	SLETIGQELKLKK
☑	108	727.7670	2180.2792	2180.1048	0.1744	1	24	1.7e+02	1	MANDTALANNLLPYVRNFK + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|261339072** Score: **195**
YfaZ family protein [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\280808\SSP 8205_RJ23_01_550.d\SSP 8205_RJ23_01_550.mgf

Nominal mass (M_r): **18559**; Calculated pI value: **6.82**
NCBI BLAST search of gi|261339072 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **24%**

Matched peptides shown in **Bold Red**

```
  1 MKKLNVLLLS ALTVVSGSAL AMGGSIEQGK NFTNLNLEMG KSSSGLYAES
 51 NWLKNTDDGT QTGGVGAGYN LEVGPVMLNA GAKAIYIGPK KGDNGVAFPI
101 GGGVNVALTD SIHVFGEGYV APDGLNNSVK NYVEANGGVS WTPIKPVTLK
151 VGYRHVSVDG KDGRPNHTLI DGAYVGGGVS F
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
31 - 41	648.8060	1295.5974	1295.6180	-0.0205	0	K.NFTNLNLEMGK.S Oxidation (M) (Ions score 55)
42 - 54	721.3580	1440.7014	1440.6885	0.0130	0	K.SSSGLYAESNWLK.N (Ions score 61)
162 - 181	678.2780	2031.8122	2030.9810	0.8311	0	K.DGRPNHTLIDGAYVGGGVSF.- (Ions score 79)

Error: try setting browser cache to automatic. Error: try setting browser cache to automatic.

```
LOCUS       ZP_05966930              181 aa            linear   BCT 15-OCT-2009
DEFINITION  YfaZ family protein [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05966930
VERSION     ZP_05966930.1  GI:261339072
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000004.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 181)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 181)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000004.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
```

```
          Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
          gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
          cancerogenus, not the sequenced strain), is a member of the
          Proteobacteria division of the domain bacteria and has been
          isolated from human feces. The sequenced strain was obtained from
          ATCC (ATCC 35316).

          We have performed one round of automated sequence improvement
          (pre-finishing), along with manual improvement that includes
          breaking apart any mis-assembly, and making manual joins where
          possible. Manual edits also are made where the consensus appears to
          be incorrect. All low quality data on the ends of contigs is
          removed. Contigs are ordered and oriented where possible.

          This work was supported by the National Human Genome Research
          Institute and the NIH Roadmap Human Microbiome Project
          [http://nihroadmap.nih.gov/hmp/].

          This is a reference genome for the Human Microbiome Project.
          This project is co-owned with the Human Microbiome Project DACC.
          Genome Coverage: 29.7x
          Sequencing Technology: 454.
          Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..181
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..181
                     /product="YfaZ family protein"
                     /calculated_mol_wt=18440
     Region          1..181
                     /region_name="YfaZ"
                     /note="YfaZ precursor; pfam07437"
                     /db_xref="CDD:116058"
     CDS             1..181
                     /locus_tag="EcanA3_020100001365"
                     /coded_by="NZ_ABWM02000004.1:301716..302261"
                     /transl_table=11
                     /db_xref="CDD:116058"
```

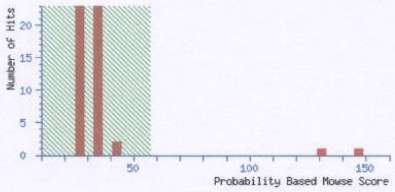
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 8206
MS data file    : D:\Data\Lakshmy\120808\241108\SSP 8206_RB19_01_769.d\SSP 8206_RB19_01_769.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 5 Jan 2010 at 12:14:59 GMT
Protein hits    : gi|237732258 ecotin [Citrobacter sp. 30_2]
                  gi|50582463  oxidoreductase [Exiguobacterium acetylicum]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|237732258 **Mass:** 19051 **Score:** 147 **Queries matched:** 5 **emPAI:** 0.63
ecotin [Citrobacter sp. 30_2]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	111	529.2380	1584.6922	1584.8763	-0.1841	0	(33)	12	1	K.LPIVVYTPENVDVK.Y
☑	113	793.3830	1584.7514	1584.8763	-0.1249	0	60	0.026	1	K.LPIVVYTPENVDVK.Y
☑	114	793.3980	1584.7814	1584.8763	-0.0949	0	(57)	0.057	1	K.LPIVVYTPENVDVK.Y
☑	115	793.4280	1584.8414	1584.8763	-0.0349	0	(39)	3.3	1	K.LPIVVYTPENVDVK.Y
☑	147	878.2990	1754.5834	1754.7828	-0.1994	0	86	4.4e-05	1	K.TLEGWGYDYYVFDK.V

Proteins matching the same set of peptides:
gi|261340622 **Mass:** 18823 **Score:** 147 **Queries matched:** 5
ecotin [Enterobacter cancerogenus ATCC 35316]

2. gi|50582463 **Mass:** 27360 **Score:** 129 **Queries matched:** 2 **emPAI:** 0.26
oxidoreductase [Exiguobacterium acetylicum]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	138	859.3720	1716.7294	1716.8570	-0.1276	0	87	5.1e-05	1	K.YTVITGASSGIGYETAK.L
☑	198	1015.4730	2028.9314	2028.9752	-0.0438	0	42	1.5	1	K.SVDLADNQNVHDLYEGLK.E

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	68	455.0960	1362.2662	1362.6204	-0.3543	0	43	1.7	1	FEDTPLSDFHR
☑	194	675.2970	2022.8692	2023.0520	-0.1829	1	40	2.1	1	VELLIGQTLKVDCNQHR
☑	178	644.9340	1931.7802	1932.0720	-0.2919	0	39	2.6	1	IDLLLATPEGWVLIDHK
☑	117	800.8320	1599.6494	1599.7927	-0.1432	0	36	6	1	GDPALPATTMEGIGVR + Oxidation (M)
☑	93	681.7720	1361.5294	1362.6568	-1.1274	0	36	13	1	HGYAFNELDLGK
☑	200	680.1840	2037.5302	2037.0565	0.4737	1	35	4.8	1	GASVRDIMGGLDGFLSISVK + Oxidation (M)
☑	67	453.2220	904.4294	905.4389	-1.0094	1	34	27	1	MQEAAGRK + Oxidation (M)
☑	124	546.8830	1637.6272	1637.7937	-0.1665	1	33	9.9	1	KADEYPITFGDGTPK
☑	101	714.5440	1427.0734	1427.7191	-0.6457	0	33	20	1	ADVFVAQLSGGMQR
☑	140	864.8670	1727.7194	1727.9604	-0.2409	1	32	16	1	INLKIPAGIESGMTLR + Oxidation (M)
☑	225	747.0320	2238.0742	2239.1174	-1.0432	1	31	20	1	QGPNLVGYETQNWFLKQR
☑	81	570.6700	1139.3254	1139.5459	-0.2204	1	30	41	1	DSSKGDVTGFK
☑	208	1037.5310	3109.5712	3108.5332	1.0380	1	30	45	1	DRITAVLMPGSVVSIANDSERPQYMATR + 2 Oxidation (M)
☑	95	692.2590	1382.5034	1382.8245	-0.3211	2	30	25	1	LSIALGKGSKNPAK
☑	219	737.6540	2209.9402	2209.1062	0.8340	2	28	36	1	NSSNIFQIKKFHTNQTMR + Oxidation (M)

```
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..167
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..167
                     /product="ecotin"
                     /calculated_mol_wt=18529
     Region          30..165
                     /region_name="Ecotin"
                     /note="Protease Inhibitor Ecotin; homodimeric protease
                     inhibitor which binds two chymotrypsin-like serine
                     proteases to form a heterotetramer. Found in bacterial
                     periplasm. Inhibits a broad range of serine proteases
                     including collagenase, trypsin, chymotrypsin...; cd00242"
                     /db_xref="CDD:29318"
     Site            order(91..95,133..138)
                     /site_type="other"
                     /note="secondary substrate binding site"
                     /db_xref="CDD:29318"
     Site            107..110
                     /site_type="other"
                     /note="primary substrate binding site"
                     /db_xref="CDD:29318"
     Site            109..110
                     /site_type="inhibition"
                     /note="inhibition loop"
                     /db_xref="CDD:29318"
     Site            155..165
                     /site_type="other"
                     /note="dimerization interface"
                     /db_xref="CDD:29318"
     CDS             1..167
                     /locus_tag="EcanA3_020100009153"
                     /coded_by="NZ_ABWM02000010.1:67048..67551"
                     /note="COG4574 Serine protease inhibitor ecotin"
                     /transl_table=11
                     /db_xref="CDD:134928"
```

Mascot: http://www.matrixscience.com/

Mascot Search Results

Protein View

Match to: **gi|261340622** Score: **147**
ecotin [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\241108\SSP 8206_RB19_01_769.d\SSP 8206_RB19_01_769.mgf

Nominal mass (M_r): **18823**; Calculated pI value: **8.33**
NCBI BLAST search of gi|261340622 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **16%**

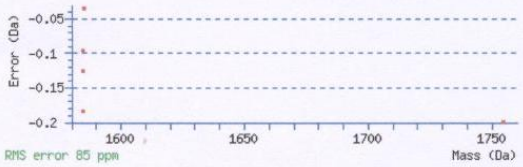
Matched peptides shown in **Bold Red**

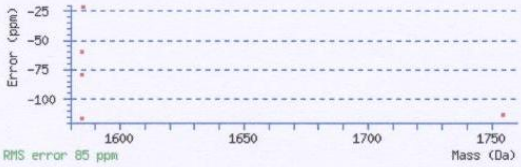
```
  1 MKNAPKFAIA LLAAACVSTS AFATDTKNDQ PLEKVAPYPQ AEKGMKRQVI
 51 QVPAQENEAN FKVELLIGQT LEVDCNHHRL GGKLESKTLE GWGYDYYVFD
101 KVTSPVSTMM ACPDGKKEKK FVTAYLGDNS LLRYNSKLPI VVYTPENVDV
151 KYRIWKADEN VGQAVIR
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence	
88 - 101	878.2990	1754.5834	1754.7828	-0.1994	0	K.TLEGWGYDYYVFDK.V	(Ions score 86)
138 - 151	529.2380	1584.6922	1584.8763	-0.1841	0	K.LPIVVYTPENVDVK.Y	(Ions score 33)
138 - 151	793.3830	1584.7514	1584.8763	-0.1249	0	K.LPIVVYTPENVDVK.Y	(Ions score 60)
138 - 151	793.3980	1584.7814	1584.8763	-0.0949	0	K.LPIVVYTPENVDVK.Y	(Ions score 57)
138 - 151	793.4280	1584.8414	1584.8763	-0.0349	0	K.LPIVVYTPENVDVK.Y	(Ions score 39)





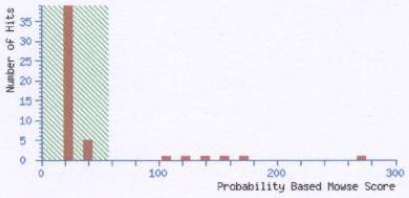

```
LOCUS       ZP_05968480              167 aa            linear   BCT 15-OCT-2009
DEFINITION  ecotin [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05968480
VERSION     ZP_05968480.1  GI:261340622
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000010.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 167)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 167)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000010.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
```

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 8707
MS data file    : D:\Data\Lakshmy\120808\280808\SSP 8707_RK17_01_587.d\SSP 8707_RK17_01_587.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 5 Jan 2010 at 12:23:26 GMT
Protein hits    : gi|261341330 translocation protein TolB [Enterobacter cancerogenus ATCC 35316]
                  gi|10120890  Chain A, Crystal Structure Of The E. Coli Tolb Protein
                  gi|62179322  translocation protein TolB [Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67]
                  gi|157369522 translocation protein TolB [Serratia proteamaculans 568]
                  gi|152969308 translocation protein TolB [Klebsiella pneumoniae subsp. pneumoniae MGH 78578]
                  gi|242238602 Tol-Pal system beta propeller repeat protein TolB [Dickeya dadantii Ech703]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected | ☐ Error tolerant

1. gi|261341330 **Mass:** 45999 **Score:** 271 **Queries matched:** 7 **emPAI:** 0.32
translocation protein TolB [Enterobacter cancerogenus ATCC 35316]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	50	571.7910	1141.5674	1141.5768	-0.0093	0	70	0.0028	1	K.LAYVTFESGR.S
☑	51	571.7940	1141.5734	1141.5768	-0.0033	0	(70)	0.0035	1	K.LAYVTFESGR.S
☑	62	714.9110	1427.8074	1427.8096	-0.0022	0	40	3.1	1	R.SALVIQTLSNGAVR.Q
☑	49	567.5860	1699.7362	1699.7876	-0.0514	0	(34)	18	1	R.SPQPLMSPAWSPDGSK.L + Oxidation (M)
☑	89	850.9190	1699.8234	1699.7591	0.0644	0	37	5.5	1	R.VSDYDGYNQFTVHR.S
☑	90	850.9250	1699.8354	1699.7876	0.0479	0	69	0.0033	1	R.SPQPLMSPAWSPDGSK.L + Oxidation (M)
☑	105	912.4500	1822.8854	1822.9247	-0.0393	0	59	0.036	1	K.TGSLNLYVMDIGSGQIR.Q

---

2. gi|10120890 **Mass:** 42745 **Score:** 170 **Queries matched:** 4 **emPAI:** 0.16
Chain A, Crystal Structure Of The E. Coli Tolb Protein
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
50	571.7910	1141.5674	1141.5768	-0.0093	0	70	0.0028	1	K.LAYVTFESGR.S
51	571.7940	1141.5734	1141.5768	-0.0033	0	(70)	0.0035	1	K.LAYVTFESGR.S
49	567.5860	1699.7362	1699.8206	-0.0844	0	34	18	1	R.SPQPLFSPAWSPDGSK.L
90	850.9250	1699.8354	1699.7876	0.0479	0	69	0.0033	1	R.SPQPLMSPAWSPDGSK.L + Oxidation (M)

---

3. gi|62179322 **Mass:** 46150 **Score:** 164 **Queries matched:** 5 **emPAI:** 0.15
translocation protein TolB [Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
50	571.7910	1141.5674	1141.5768	-0.0093	0	70	0.0028	1	K.LAYVTFESGR.S
51	571.7940	1141.5734	1141.5768	-0.0033	0	(70)	0.0035	1	K.LAYVTFESGR.S
62	714.9110	1427.8074	1427.8096	-0.0022	0	24	1.2e+02	3	R.SALVIQTLANGSVR.Q
49	567.5860	1699.7362	1699.7876	-0.0514	0	(34)	18	1	R.SPQPLMSPAWSPDGSK.L + Oxidation (M)
90	850.9250	1699.8354	1699.7876	0.0479	0	69	0.0033	1	R.SPQPLMSPAWSPDGSK.L + Oxidation (M)

Proteins matching the same set of peptides:
gi|224582567 **Mass:** 46584 **Score:** 164 **Queries matched:** 5
translocation protein TolB precursor [Salmonella enterica subsp. enterica serovar Paratyphi C strain RKS4594]

Mascot Search Results

Protein View

Match to: **gi|261341330** Score: **271**
translocation protein TolB [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\280808\SSP 8707_RK17_01_587.d\SSP 8707_RK17_01_587.mgf

Nominal mass (M_r): **45999**; Calculated pI value: **6.98**
NCBI BLAST search of gi|261341330 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **16%**

Matched peptides shown in **Bold Red**

```
  1 MMKQALRVAF GFLMLWAAVL HAEVRIEITQ GVDSARPIGV VPFQWAGPGA
 51 APEDIGGIVA ADLRNSGKFN PLDRSRLPQQ PGTAQEVQPA AWSALGIDAV
101 VVGQVTPAPD GGYNVAYQLV DTGGAPGTVL AQNTYKVNKQ WLRYAGHTAS
151 DEVFEKLTGI KGAFRTRIAY VVQTNGGQFP YELRVSDYDG YNQFTVHRSP
201 QPLMSPAWSP DGSKLAYVTF ESGRSALVIQ TLSNGAVRQV ASFPRHNGAP
251 AFSPDGTKLA FALSKTGSLN LYVMDIGSGQ IRQITDGRSN NTEPTWFPDS
301 QNLAFTSDQA GRPQVYKVNI NGGAAQRITW EGSQNQDADI SADGKTMVMV
351 STAGGQQHIA KQDLVTGGVQ VLSSTFLDET PSLAPNGTMV IYSSSQGMGS
401 VLNLVSTDGR FKARIPATDG QVKSPAWSPY L
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
185 - 198	850.9190	1699.8234	1699.7591	0.0644	0	R.VSDYDGYNQFTVHR.S (Ions score 37)
199 - 214	567.5860	1699.7362	1699.7876	-0.0514	0	R.SPQPLMSPAWSPDGSK.L Oxidation (M) (Ions score 34)
199 - 214	850.9250	1699.8354	1699.7876	0.0479	0	R.SPQPLMSPAWSPDGSK.L Oxidation (M) (Ions score 69)
215 - 224	571.7910	1141.5674	1141.5768	-0.0093	0	K.LAYVTFESGR.S (Ions score 70)
215 - 224	571.7940	1141.5734	1141.5768	-0.0033	0	K.LAYVTFESGR.S (Ions score 70)
225 - 238	714.9110	1427.8074	1427.8096	-0.0022	0	R.SALVIQTLSNGAVR.Q (Ions score 40)
266 - 282	912.4500	1822.8854	1822.9247	-0.0393	0	K.TGSLNLYVMDIGSGQIR.Q (Ions score 59)

Error: try setting browser cache to automatic. Error: try setting browser cache to automatic.

```
LOCUS       ZP_05969188              431 aa            linear   BCT 15-OCT-2009
DEFINITION  translocation protein TolB [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05969188
VERSION     ZP_05969188.1  GI:261341330
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000020.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 431)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 431)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000020.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
```

```
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..431
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..431
                     /product="translocation protein TolB"
                     /calculated_mol_wt=45897
     Region          2..431
                     /region_name="tolB"
                     /note="translocation protein TolB; Provisional; PRK03629"
                     /db_xref="CDD:134902"
     Region          31..174
                     /region_name="TolB_N"
                     /note="TolB amino-terminal domain; pfam04052"
                     /db_xref="CDD:112850"
     Region          238..273
                     /region_name="PD40"
                     /note="WD40-like Beta Propeller Repeat; pfam07676"
                     /db_xref="CDD:116290"
     Region          282..317
                     /region_name="PD40"
                     /note="WD40-like Beta Propeller Repeat; pfam07676"
                     /db_xref="CDD:116290"
     CDS             1..431
                     /gene="tolB"
                     /locus_tag="EcanA3_020100012746"
                     /coded_by="NZ_ABWM02000020.1:251385..252680"
                     /note="COG0823 Periplasmic component of the Tol biopolymer
                     transport system"
                     /transl_table=11
                     /db_xref="CDD:134902"
```

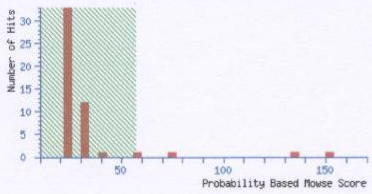
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 9601
MS data file    : D:\Data\Lakshmy\120808\140808\SSP9601_RF13_01_334.d\SSP9601_RF13_01_334.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 5 Jan 2010 at 12:24:38 GMT
Protein hits    : gi|148368     outer membrane protein II [Enterobacter aerogenes]
                  gi|146311681 acetoin reductase [Enterobacter sp. 638]
                  gi|157146352 outer membrane protein A [Citrobacter koseri ATCC BAA-895]
                  gi|146312800 glycine betaine transporter periplasmic subunit [Enterobacter sp. 638]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|148368 **Mass:** 25654 **Score:** 152 **Queries matched:** 3 **emPAI:** 0.28
outer membrane protein II [Enterobacter aerogenes]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	90	632.8000	1263.5854	1263.6460	-0.0605	0	41	3.7	2	K.DGSVVVLGFTDR.I
☑	112	813.9580	1625.9014	1625.7937	0.1077	0	42	1.7	1	K.LGYPVTDDLDVYTR.L
☑	235	783.3450	2347.0132	2346.1856	0.8276	0	69	0.0027	1	R.FGQQEDAPVVAPAPAPAPEVQTK.H

Proteins matching the same set of peptides:
gi|261339284 **Mass:** 37561 **Score:** 152 **Queries matched:** 3
outer membrane protein A [Enterobacter cancerogenus ATCC 35316]

2. gi|146311681 **Mass:** 26942 **Score:** 137 **Queries matched:** 3 **emPAI:** 0.26
acetoin reductase [Enterobacter sp. 638]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	126	564.6150	1690.8232	1689.8686	0.9546	1	20	2.5e+02	4	K.VDVSNREQVFAAVEK.A
☑	132	857.3840	1712.7534	1712.7893	-0.0359	0	55	0.085	1	K.DGFAVAIADYNEETAK.A
☑	238	802.0750	2403.2032	2404.1329	-0.9297	0	61	0.017	1	K.IINACSQAGHTGNPELAVYSSSK.F

Proteins matching the same set of peptides:
gi|261341164 **Mass:** 26884 **Score:** 137 **Queries matched:** 3
acetoin reductase [Enterobacter cancerogenus ATCC 35316]

3. gi|157146352 **Score:** 77 **Queries matched:** 2
outer membrane protein A [Citrobacter koseri ATCC BAA-895]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
90	632.8000	1263.5854	1263.6460	-0.0605	0	41	3.7	2	K.DGSVVVLGFTDR.I
235	783.3450	2347.0132	2346.1855	0.8276	0	36	5.2	2	R.FGQQEEAAPIAPAPAPAPEVQTK.H

4. gi|146312800 **Mass:** 36495 **Score:** 58 **Queries matched:** 1 **emPAI:** 0.09
glycine betaine transporter periplasmic subunit [Enterobacter sp. 638]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	119	835.3920	1668.7694	1668.8359	-0.0665	0	58	0.043	1	R.EGTFVTGAAQGYLIDK.K

Mascot Search Results

Protein View

Match to: **gi|261339284** Score: **152**
outer membrane protein A [Enterobacter cancerogenus ATCC 35316]
Found in search of D:\Data\Lakshmy\120808\140808\SSP9601_RF13_01_334.d\SSP9601_RF13_01_334.mgf

Nominal mass (M_r): **37561**; Calculated pI value: **5.19**
NCBI BLAST search of gi|261339284 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Enterobacter cancerogenus ATCC 35316

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **14%**

Matched peptides shown in **Bold Red**

```
  1 MKKTAIAIAV ALAGFATVAQ AAPKDNTWYA GGKLGWSQFH DTGWYNSSLN
 51 NDGPTHESQL GAGAFGGYQV NPYVGFEMGY DWLGRMPYKG DNVNGAFKAQ
101 GVQLTAKLGY FVTDDLDVYT RLGGMVWRAD SSNSIAGDDH DTGVSPVFAG
151 GVEWAMTRDI ATRLEYQWVN NIGDGATVGV RPDNGMLSVG VSYRFGQQED
201 APVVAPAPAP APEVQTKHFT LKSDVLFNFN KATLKPEGQQ ALDQLYTQLS
251 NLDPKDGSVV VLGFTDRIGS DAYNQGLSEK RAQSVVDYLV SKGIPANKIS
301 PRGMGESNPV TGNTCDNVKP RAALIDCLAP DRRVEIEVKG IKDVVTQPAA
351
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
108 - 121	813.9580	1625.9014	1625.7937	0.1077	0	K.LGYPVTDDLDVYTR.L (Ions score 42)
195 - 217	783.3450	2347.0132	2346.1856	0.8276	0	R.FGQQEDAPVVAPAPAPAPEVQTK.H (Ions score 69)
256 - 267	632.8000	1263.5854	1263.6460	-0.0605	0	K.DGSVVVLGFTDR.I (Ions score 41)

Error: try setting browser cache to automatic. Error: try setting browser cache to automatic.

```
LOCUS       ZP_05967142              350 aa            linear   BCT 15-OCT-2009
DEFINITION  outer membrane protein A [Enterobacter cancerogenus ATCC 35316].
ACCESSION   ZP_05967142
VERSION     ZP_05967142.1  GI:261339284
DBLINK      Project:28663
DBSOURCE    REFSEQ: accession NZ_ABWM02000004.1
KEYWORDS    .
SOURCE      Enterobacter cancerogenus ATCC 35316
  ORGANISM  Enterobacter cancerogenus ATCC 35316
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Enterobacter.
REFERENCE   1  (residues 1 to 350)
  AUTHORS   Weinstock,G., Sodergren,E., Clifton,S., Fulton,L., Fulton,B.,
            Courtney,L., Fronick,C., Harrison,M., Strong,C., Farmer,C.,
            Delahaunty,K., Markovic,C., Hall,O., Minx,P., Tomlinson,C.,
            Mitreva,M., Nelson,J., Hou,S., Wollam,A., Pepin,K.H., Johnson,M.,
            Bhonagiri,V., Nash,W.E., Warren,W., Chinwalla,A., Mardis,E.R. and
            Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2009) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
REFERENCE   2  (residues 1 to 350)
  AUTHORS   Fulton,L., Clifton,S., Fulton,B., Xu,J., Minx,P., Pepin,K.H.,
            Johnson,M., Thiruvilangam,P., Bhonagiri,V., Nash,W.E., Mardis,E.R.
            and Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2008) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park, St. Louis, MO
            63108, USA
COMMENT     WGS REFSEQ: This record is provided to represent a collection of
            whole genome shotgun sequences. The reference sequence was derived
            from ABWM02000004.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
```

```
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
            Enterobacter cancerogenus (GenBank Accession Number for 16S rDNA
            gene: Z96078, the 16S rDNA gene of a related strain of Enterobacter
            cancerogenus, not the sequenced strain), is a member of the
            Proteobacteria division of the domain bacteria and has been
            isolated from human feces. The sequenced strain was obtained from
            ATCC (ATCC 35316).

            We have performed one round of automated sequence improvement
            (pre-finishing), along with manual improvement that includes
            breaking apart any mis-assembly, and making manual joins where
            possible. Manual edits also are made where the consensus appears to
            be incorrect. All low quality data on the ends of contigs is
            removed. Contigs are ordered and oriented where possible.

            This work was supported by the National Human Genome Research
            Institute and the NIH Roadmap Human Microbiome Project
            [http://nihroadmap.nih.gov/hmp/].

            This is a reference genome for the Human Microbiome Project.
            This project is co-owned with the Human Microbiome Project DACC.
            Genome Coverage: 29.7x
            Sequencing Technology: 454.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..350
                     /organism="Enterobacter cancerogenus ATCC 35316"
                     /strain="ATCC 35316"
                     /db_xref="taxon:500639"
     Protein         1..350
                     /product="outer membrane protein A"
                     /calculated_mol_wt=37337
     Region          1..346
                     /region_name="PRK10808"
                     /note="outer membrane protein A; Reviewed; PRK10808"
                     /db_xref="CDD:138188"
     Region          23..196
                     /region_name="Surface_Ag_2"
                     /note="Surface antigen; c101155"
                     /db_xref="CDD:141128"
     Region          224..338
                     /region_name="OmpA_C-like"
                     /note="Peptidoglycan binding domains similar to the
                     C-terminal domain of outer-membrane protein OmpA; cd07185"
                     /db_xref="CDD:143586"
     Site            order(229..230,265..266,269,273..274,277,328,332)
                     /site_type="other"
                     /note="ligand binding site"
                     /db_xref="CDD:143586"
     CDS             1..350
                     /locus_tag="EcanA3_020100002425"
                     /coded_by="complement(NZ_ABWM02000004.1:545400..546452)"
                     /note="COG2885 Outer membrane protein and related
                     peptidoglycan-associated (lipo)proteins"
                     /transl_table=11
                     /db_xref="CDD:138188"
```

Mascot: http://www.matrixscience.com/

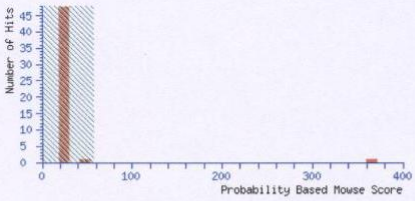
G3 Mass spectrometric analysis data for differentially expressed spots *B. fragilis*

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 1104
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (1104) 10_RJ17_01_1396.d\SSP (1104) 10_RJ17_01_1396.mgf
Database        : NCBInr 20100108 (10291680 sequences; 3511877860 residues)
Taxonomy        : Bacteria (Eubacteria) (5718488 sequences)
Timestamp       : 13 Jan 2010 at 10:39:58 GMT
Protein hits    : gi|53713703 TPR repeat-containing protein [Bacteroides fragilis YCH46]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|53713703 **Mass:** 46634 **Score:** 366 **Queries matched:** 8 **emPAI:** 0.51
TPR repeat-containing protein [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	125	776.2740	1550.5334	1550.7365	-0.2031	0	64	0.0073	1	K.DNAATWDVAGYIQK.R
☑	140	835.8360	1669.6574	1669.8522	-0.1948	0	74	0.00086	1	K.AEQLINEALTNPETK.D
☑	185	663.5640	1987.6702	1986.9357	* 0.7345	1	51	0.13	1	K.DKENGQFAMQLLTDAYK.A + Oxidation (M)
☑	186	994.8670	1987.7194	1986.9357	0.7838	1	(26)	97	1	K.DKENGQFAMQLLTDAYK.A + Oxidation (M)
☑	187	995.8420	1989.6694	1989.9030	-0.2335	0	68	0.0026	1	R.VYYNLNMGPEFEEIEK.M + Oxidation (M)
	233	**762.6100**	**2284.8082**	**2283.9738**	**0.8344**	**1**	**19**	**2.1e+02**	**2**	**R.VYYNLNMGPEFEEIEKMM.- + 3 Oxidation (M)**
☑	238	797.0170	2388.0292	2387.2961	0.7331	0	32	12	1	K.TILAERPNLINGGIQYFNLNK.N
☑	248	930.6050	2788.7932	2789.2245	-0.4313	0	57	0.03	1	K.FPENQYFFANLVDYYSSSNQNDK.A

Proteins matching the same set of peptides:
gi|60681974 **Mass:** 45773 **Score:** 366 **Queries matched:** 8
hypothetical protein BF2494 [Bacteroides fragilis NCTC 9343]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	221	737.9210	2210.7412	2210.1115	0.6297	0	34	6.3	1	MYILGINSVYHESAVCLLK
☑	131	804.8580	1607.7014	1607.8125	-0.1110	0	33	14	1	MPIYNELMLMPLK + Oxidation (M)
☑	234	766.9800	2297.9182	2299.0790	-1.1609	2	32	11	1	LNGYWEIDKMRSLTDDAEK + Oxidation (M)
☑	80	629.4150	1885.2232	1885.9707	-0.7475	0	32	32	1	SIMDNLGLPVIDDEIVK + Oxidation (M)
☑	121	752.3550	2254.0432	2255.0205	-0.9773	1	30	51	1	LNEPDFDGDDWKIEMFLR + Oxidation (M)
☑	225	1113.0610	2224.1074	2224.0437	0.0638	0	30	25	1	DISWGSTWGAGDNVDIYQLK
☑	44	391.1050	1170.2932	1169.6227	0.6705	0	28	59	1	LLAMEGVNPAR
☑	88	650.7300	1299.4454	1298.7921	0.6533	2	27	44	1	KNESLKLVLQK
☑	157	585.0580	1752.1522	1752.9312	-0.7790	2	27	40	1	VIGGNFKRFQFTPDK
☑	108	707.5460	1413.0774	1412.7848	0.2927	2	27	77	1	VAVATEEGRRVAR
☑	79	623.1710	1866.4912	1865.8263	0.6649	1	26	91	1	LWGGYAKGMTAMDYMR + Oxidation (M)
☑	229	747.8320	2240.4742	2240.2277	0.2465	1	26	40	1	YQVDLLTLHATAGRQALTAAK
☑	126	779.9360	2336.7862	2336.1722	0.6140	0	26	1.1e+02	1	GYEVFLINALSGEGLPELMER
☑	209	700.9180	2099.7322	2099.9941	-0.2620	0	25	49	1	TINDMQMYIECVQLVVK + Oxidation (M)
☑	196	681.0290	2040.0652	2040.9350	-0.8698	2	25	74	1	VEEIPDMYGWKKETAET + Oxidation (M)
☑	145	563.1340	1686.3802	1685.9365	0.4436	2	25	56	1	AGLIHAFFNDRKVAK
☑	112	721.6110	2161.8112	2162.0578	-0.2466	2	25	1.1e+02	1	ATYSDEFYMARALELARR
☑	204	694.9280	2081.7622	2081.9795	-0.2174	1	25	55	1	MAEMVIGEKVTLEEMGGAR + 2 Oxidation (M)
☑	142	838.0410	2511.1012	2511.2329	-0.1317	1	25	1.4e+02	1	SVIYEMHVGGFTRHPNSGVAPEK
☑	213	711.0680	2130.1822	2129.9721	0.2101	2	25	82	1	CDTEEIRKYAICSLSQR

Mascot Search Results

Protein View

Match to: **gi|60681974** Score: **366**
hypothetical protein BF2494 [Bacteroides fragilis NCTC 9343]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (1104) 10_RJ17_01_1396.d\SSP (1104) 10_RJ17_01_1396.mgf

Nominal mass (M_r): **45773**; Calculated pI value: **5.24**
NCBI BLAST search of gi|60681974 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis NCTC 9343
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60493408 from Bacteroides fragilis NCTC 9343

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **27%**

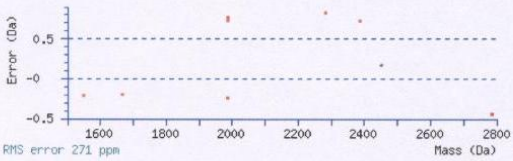
Matched peptides shown in Bold Red

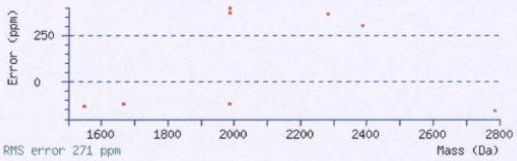
```
  1 MVLLMAVSFA FAQEKNVKEA KSIAGEVKPD FAKAEQLINE ALTNPETKDN
 51 AATWDVAGYI QKRINEKEME NAYLRKPYDT LKVYNSVLNM YNYYVKCDEL
101 AQIPNEKGKI KNKYRSANSK TILAERPNLI NGGIQYFNLN KNEDALKYFA
151 AYVDAATLPM MEKENLLEKD TILPQVAYYA TLAADRVGDK DAVMKYAQYA
201 LKDKENGQFA MQLLTDAYKA KGDTAKWVEK LQEGIVKFPE NQYFFANLVD
251 YYSSSNQNDK AMQFADDMLA KDPNNKLYLY VKAYLYHNMK DYEKAIEFYK
301 KTLDIDPAYA EACSNLGLVY LLQAQEYADK APADINDPNY ATAQAEIKKF
351 YEAAKPYYEK ARELKPDQKD LWLQGLYRVY YNLNMGPEFE EIEKMM
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
34 - 48	835.8360	1669.6574	1669.8522	-0.1948	0	K.AEQLINEALTNPETK.D (Ions score 74)
49 - 62	776.2740	1550.5334	1550.7365	-0.2031	0	K.DNAATWDVAGYIQK.R (Ions score 64)
121 - 141	797.0170	2388.0292	2387.2961	0.7331	0	K.TILAERPNLINGGIQYFNLNK.N (Ions score 32)
203 - 219	663.5640	1987.6702	1986.9357	0.7345	1	K.DKENGQFAMQLLTDAYK.A Oxidation (M) (Ions score 51)
203 - 219	994.8670	1987.7194	1986.9357	0.7838	1	K.DKENGQFAMQLLTDAYK.A Oxidation (M) (Ions score 26)
238 - 260	930.6050	2788.7932	2789.2245	-0.4313	0	K.FPENQYFFANLVDYYSSSNQNDK.A (Ions score 57)
379 - 394	995.8420	1989.6694	1989.9030	-0.2335	0	R.VYYNLNMGPEFEEIEK.M Oxidation (M) (Ions score 68)
379 - 396	762.6100	2284.8082	2283.9738	0.8344	1	R.VYYNLNMGPEFEEIEKMM.- 3 Oxidation (M) (Ions score 19)






```
LOCUS       YP_212118                396 aa            linear   BCT 01-MAY-2009
DEFINITION  hypothetical protein BF2494 [Bacteroides fragilis NCTC 9343].
ACCESSION   YP_212118
VERSION     YP_212118.1  GI:60681974
DBLINK      Project:46
DBSOURCE    REFSEQ: accession NC_003228.3
KEYWORDS    .
SOURCE      Bacteroides fragilis NCTC 9343
  ORGANISM  Bacteroides fragilis NCTC 9343
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 396)
  AUTHORS   Cerdeno-Tarraga,A.M., Patrick,S., Crossman,L.C., Blakely,G.,
            Abratt,V., Lennard,N., Poxton,I., Duerden,B., Harris,B.,
            Quail,M.A., Barron,A., Clark,L., Corton,C., Doggett,J.,
            Holden,M.T., Larke,N., Line,A., Lord,A., Norbertczak,H., Ormond,D.,
            Price,C., Rabbinowitsch,E., Woodward,J., Barrell,B. and Parkhill,J.
  TITLE     Extensive DNA inversions in the B. fragilis genome control variable
            gene expression
  JOURNAL   Science 307 (5714), 1463-1465 (2005)
   PUBMED   15746427
REFERENCE   2  (residues 1 to 396)
  AUTHORS   Cerdeno-Tarraga,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2004) Pathogen Sequencing Unit, Sanger Institute,
            Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United
            Kingdom
REFERENCE   3  (residues 1 to 396)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
```

```
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from CAH08194.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..396
                     /organism="Bacteroides fragilis NCTC 9343"
                     /strain="ATCC 25285; NCTC 9343"
                     /db_xref="ATCC:25285"
                     /db_xref="taxon:272559"
     Protein         1..396
                     /product="hypothetical protein"
                     /calculated_mol_wt=45555
     Region          <173..>373
                     /region_name="PEP_TPR_lipo"
                     /note="putative PEP-CTERM system TPR-repeat lipoprotein;
                     TIGR02917"
                     /db_xref="CDD:131963"
     Region          247..323
                     /region_name="TPR"
                     /note="Tetratricopeptide repeat domain; typically contains
                     34 amino acids
                     [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-
                     X(2)-[ASL]-X(4)-[PKE] is the consensus sequence; found in
                     a variety of organisms including bacteria, cyanobacteria,
                     yeast, fungi...; cd00189"
                     /db_xref="CDD:29151"
     Site            order(247,250..251,280..281,283..284,287..288,314..315,
                     318..319,322..323)
                     /site_type="binding"
                     /note="binding surface"
                     /db_xref="CDD:29151"
     Site            order(249,261,265,268,283,295,299,302,317)
                     /site_type="other"
                     /note="TPR motif"
                     /db_xref="CDD:29151"
     CDS             1..396
                     /locus_tag="BF2494"
                     /coded_by="NC_003228.3:2901913..2903103"
                     /note="Similar to Bacteroides thetaiotaomicron putative
                     tetratricopeptide repeat family protein BT0900
                     SWALL:Q8A9B5 (EMBL:AE016929) (398 aa) fasta scores: E():
                     4.1e-110, 72.79% id in 397 aa"
                     /transl_table=11
                     /db_xref="GeneID:3287481"
```

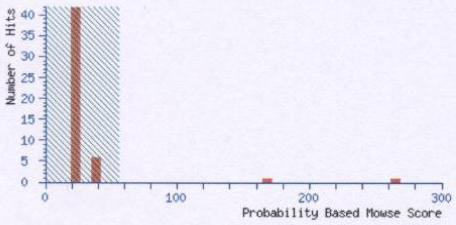
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 2101
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2101) 10_RI4_01_1304.d\SSP (2101) 10_RI4_01_1304.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:14:56 GMT
Protein hits    : gi|53715478  50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46]
                  gi|189460676 hypothetical protein BACCOP_01323 [Bacteroides coprocola DSM 17136]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 56 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53715478 **Mass:** 12688 **Score:** 265 **Queries matched:** 4 **emPAI:** 0.61
50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 60	565.2530	1128.4914	1128.6390	-0.1476	0	33	13	1	K.EVNELATILK.E
☑ 71	644.2390	1286.4634	1287.6558	-1.1923	0	52	0.14	1	K.TLEEAGAEVELK.-
☑ 72	666.8020	1331.5894	1331.7449	-0.1554	0	88	3.6e-05	1	K.AFAEQLVNLTVK.E
☑ 234	852.4230	2554.2472	2554.2550	-0.0079	0	91	1.1e-05	1	K.EEYGIEPAAAAVAVAAGPAAGAAAAEEK.S

Proteins matching the same set of peptides:
gi|150003396 **Mass:** 12704 **Score:** 265 **Queries matched:** 4
50S ribosomal protein L7/L12 [Bacteroides vulgatus ATCC 8482]
gi|153805949 **Mass:** 12686 **Score:** 265 **Queries matched:** 4
hypothetical protein BACCAC_00194 [Bacteroides caccae ATCC 43185]

---

2. gi|189460676 **Mass:** 12771 **Score:** 165 **Queries matched:** 3 **emPAI:** 0.27
hypothetical protein BACCOP_01323 [Bacteroides coprocola DSM 17136]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
60	565.2530	1128.4914	1128.6390	-0.1476	0	33	13	1	K.EVNELATILK.E
71	644.2390	1286.4634	1285.6765	0.7869	0	43	1	2	K.VLEEAGAEVELK.-
72	666.8020	1331.5894	1331.7449	-0.1554	0	88	3.6e-05	1	K.AFAEQLVNLTVK.E

Peptide matches not assigned to protein hits: (no details means no match)

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 180	725.9040	2174.6902	2175.0960	-0.4059	0	33	6.2	1	GPLTQGYLGTPQYSAPEQLR
☑ 160	1041.9470	2081.8794	2083.0334	-1.1540	2	32	9.8	1	SGYERVDIVENKGEFSVR
☑ 213	763.0340	2286.0802	2287.0361	-0.9560	1	29	20	1	SGGIGLTQAFYRAMMEAQNGGE
☑ 146	673.3010	2016.8812	2017.0844	-0.2032	0	29	21	1	ITQLYEGTNGIQALDLIR
☑ 250	970.7500	2909.2282	2908.6320	0.5962	0	29	13	1	QRPILAEGPAIILVNPQLPENIGMVAR
65	584.5490	1750.6252	1749.9050	0.7202	0	27	51	1	HLVTALFELEDGGPPR

Mascot Search Results

Protein View

Match to: **gi|53715478** Score: **265**
50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2101) 10_RI4_01_1304.d\SSP (2101) 10_RI4_01_1304.mgf

Nominal mass (M_r): **12688**; Calculated pI value: **4.71**
NCBI BLAST search of gi|53715478 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60683451 from Bacteroides fragilis NCTC 9343
gi|253566657 from Bacteroides sp. 3_2_5
gi|255011634 from Bacteroides fragilis 3_1_12
gi|265767535 from Bacteroides sp. 2_1_16
gi|81313501 from Bacteroides fragilis NCTC 9343
gi|81608231 from Bacteroides fragilis
gi|52218343 from Bacteroides fragilis YCH46
gi|60494885 from Bacteroides fragilis NCTC 9343
gi|251944829 from Bacteroides sp. 3_2_5
gi|263252840 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **50%**

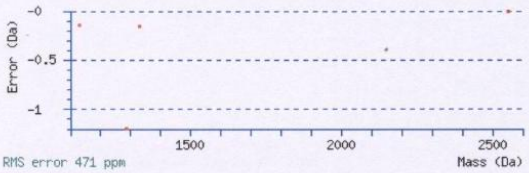
Matched peptides shown in Bold Red

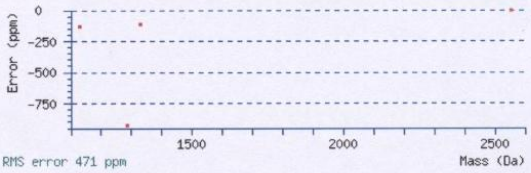
```
  1 MADLKAFAEQ LVNLTVKEVN ELATILKEEY GIEPAAAAVA VAAGPAAGAA
 51 AAEEKSSFDV VLKSAGAAKL QVVKAVKEAC GLGLKEAKDM VDGAPSVVKE
101 GLAKDEAESL KKTLEEAGAE VELK
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence	
6 - 17	666.8020	1331.5894	1331.7449	-0.1554	0	K.AFAEQLVNLTVK.E	(Ions score 88)
18 - 27	565.2530	1128.4914	1128.6390	-0.1476	0	K.EVNELATILK.E	(Ions score 33)
28 - 55	852.4230	2554.2472	2554.2550	-0.0079	0	K.EEYGIEPAAAAVAVAAGPAAGAAAAEEK.S	(Ions score 91)
113 - 124	644.2390	1286.4634	1287.6558	-1.1923	0	K.TLEEAGAEVELK.-	(Ions score 52)






```
LOCUS       YP_101470                124 aa            linear   BCT 26-APR-2009
DEFINITION  50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46].
ACCESSION   YP_101470
VERSION     YP_101470.1  GI:53715478
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 124)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 124)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 124)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from BAD50936.
```

```
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..124
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..124
                     /product="50S ribosomal protein L7/L12"
                     /calculated_mol_wt=12506
     Region          2..123
                     /region_name="Ribosomal_L7_L12"
                     /note="Ribosomal protein L7/L12. Ribosomal protein L7/L12
                     refers to the large ribosomal subunit proteins L7 and L12,
                     which are identical except that L7 is acetylated at the N
                     terminus. It is a component of the L7/L12 stalk, which is
                     located at the surface of...; cd00387"
                     /db_xref="CDD:100102"
     Site            order(2,15..17,20..21,24,27,42,45..46,48..50,56..57,78,
                     80..81,83,85..86,100..101,103,106)
                     /site_type="other"
                     /note="core dimer interface"
                     /db_xref="CDD:100102"
     Site            order(5,8,12,25,30)
                     /site_type="other"
                     /note="peripheral dimer interface"
                     /db_xref="CDD:100102"
     Site            order(19,22..23,27,30)
                     /site_type="other"
                     /note="L10 interface"
                     /db_xref="CDD:100102"
     Site            order(69..70,73..74,77,84..85,88)
                     /site_type="other"
                     /note="L11 interface"
                     /db_xref="CDD:100102"
     Site            order(70,74,77,84..85,88)
                     /site_type="other"
                     /note="putative EF-Tu interaction site"
                     /db_xref="CDD:100102"
     Site            order(70,73..74,77)
                     /site_type="other"
                     /note="putative EF-G interaction site"
                     /db_xref="CDD:100102"
     CDS             1..124
                     /gene="rplL"
                     /locus_tag="BF4193"
                     /coded_by="complement(NC_006347.1:4790804..4791178)"
                     /note="present in two forms; L12 is normal, while L7 is
                     aminoacylated at the N-terminal serine; the only multicopy
                     ribosomal protein; 4:1 ratio of L7/L12 per ribosome; two
                     L12 dimers bind L10; critically important for translation
                     efficiency and fidelity; stimulates GTPase activity of
                     translation factors"
                     /transl_table=11
                     /db_xref="GeneID:3085275"
```

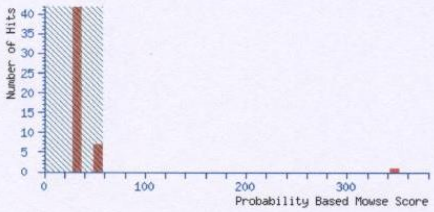
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 2102
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2102)_RE11_01_1007.d\SSP (2102)_RE11_01_1007.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:16:07 GMT
Protein hits    : gi|53715478 50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|53715478 **Mass:** 12688 **Score:** 347 **Queries matched:** 30 **emPAI:** 1.05
   50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46]
   ☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	33	447.6820	893.3494	893.4858	-0.1364	0	(35)	8.2	2	K.SSFDVVLK.S
☑	34	447.6950	893.3754	893.4858	-0.1104	0	(34)	11	1	K.SSFDVVLK.S
☑	35	447.7160	893.4174	893.4858	-0.0684	0	47	0.54	1	K.SSFDVVLK.S
	36	447.7320	893.4494	893.4858	-0.0364	0	(41)	2.2	2	K.SSFDVVLK.S
☑	37	447.7340	893.4534	893.4858	-0.0324	0	(36)	6.3	1	K.SSFDVVLK.S
☑	38	447.7350	893.4554	893.4858	-0.0304	0	(42)	1.9	1	K.SSFDVVLK.S
☑	39	447.7600	893.5054	893.4858	0.0196	0	(32)	16	1	K.SSFDVVLK.S
☑	65	565.2660	1128.5174	1128.6390	-0.1216	0	34	12	1	K.EVNELATILK.E
☑	66	567.2240	1132.4334	1132.5434	-0.1100	0	41	2.6	1	K.DMVDGAPSVVK.E + Oxidation (M)
☑	69	644.7600	1287.5054	1287.6558	-0.1503	0	(48)	0.42	1	K.TLEEAGAEVELK.-
☑	70	644.7780	1287.5414	1287.6558	-0.1143	0	(63)	0.016	1	K.TLEEAGAEVELK.-
☑	71	644.7810	1287.5474	1287.6558	-0.1083	0	(51)	0.24	1	K.TLEEAGAEVELK.-
☑	72	644.7840	1287.5534	1287.6558	-0.1023	0	(69)	0.0041	1	K.TLEEAGAEVELK.-
☑	73	644.7860	1287.5574	1287.6558	-0.0983	0	82	0.00022	1	K.TLEEAGAEVELK.-
☑	74	644.8000	1287.5854	1287.6558	-0.0703	0	(68)	0.0054	1	K.TLEEAGAEVELK.-
☑	75	644.8140	1287.6134	1287.6558	-0.0423	0	(44)	2.2	1	K.TLEEAGAEVELK.-
☑	76	644.8420	1287.6694	1287.6558	0.0137	0	(66)	0.0086	1	K.TLEEAGAEVELK.-
☑	82	444.8610	1331.5612	1331.7449	-0.1837	0	(42)	1.7	1	K.AFAEQLVNLTVK.E
☑	83	666.8020	1331.5894	1331.7449	-0.1554	0	(84)	0.00011	1	K.AFAEQLVNLTVK.E
☑	84	666.8030	1331.5914	1331.7449	-0.1534	0	97	5.7e-06	1	K.AFAEQLVNLTVK.E
☑	85	666.8080	1331.6014	1331.7449	-0.1434	0	(75)	0.00091	1	K.AFAEQLVNLTVK.E
☑	86	666.8130	1331.6114	1331.7449	-0.1334	0	(77)	0.00054	1	K.AFAEQLVNLTVK.E
☑	87	666.8320	1331.6494	1331.7449	-0.0954	0	(89)	3.8e-05	1	K.AFAEQLVNLTVK.E
☑	88	666.8340	1331.6534	1331.7449	-0.0914	0	(88)	5.1e-05	1	K.AFAEQLVNLTVK.E
☑	89	666.8400	1331.6654	1331.7449	-0.0794	0	(80)	0.00033	1	K.AFAEQLVNLTVK.E
☑	90	666.8420	1331.6694	1331.7449	-0.0754	0	(88)	4.6e-05	1	K.AFAEQLVNLTVK.E
☑	91	666.8660	1331.7174	1331.7449	-0.0274	0	(94)	1.2e-05	1	K.AFAEQLVNLTVK.E
☑	92	666.8780	1331.7414	1331.7449	-0.0034	0	(80)	0.00029	1	K.AFAEQLVNLTVK.E
☑	107	730.8450	1459.6754	1460.7181	-1.0426	1	36	7.8	1	K.EAKDMVDGAPSVVK.E + Oxidation (M)
☑	136	880.4590	1758.9034	1758.9880	-0.0845	1	12	1.6e+03	1	M.ADLKAFAEQLVNLTVK.E

Peptide matches not assigned to protein hits: (no details means no match)

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide

Mascot Search Results

Protein View

Match to: **gi|53715478** Score: **347**
50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2102)_RE11_01_1007.d\SSP (2102)_RE11_01_1007.mgf

Nominal mass (M_r): **12688**; Calculated pI value: **4.71**
NCBI BLAST search of gi|53715478 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60683451 from Bacteroides fragilis NCTC 9343
gi|253566657 from Bacteroides sp. 3_2_5
gi|255011634 from Bacteroides fragilis 3_1_12
gi|265767535 from Bacteroides sp. 2_1_16
gi|81313501 from Bacteroides fragilis NCTC 9343
gi|81608231 from Bacteroides fragilis
gi|52218343 from Bacteroides fragilis YCH46
gi|60494885 from Bacteroides fragilis NCTC 9343
gi|251944829 from Bacteroides sp. 3_2_5
gi|263252840 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **48%**

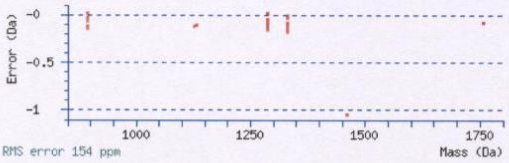
Matched peptides shown in Bold Red

```
  1 MADLKAFAEQ LVNLTVKEVN ELATILKEEY GIEPAAAAVA VAAGPAAGAA
 51 AAEEKSSFDV VLKSAGAAKL QVVKAVKEAC GLGLKEAKDM VDGAPSVVKE
101 GLAKDEAESL KKTLEEAGAE VELK
```

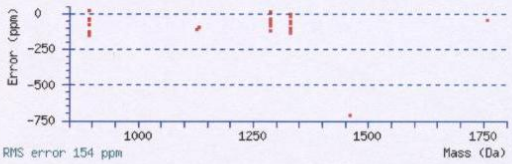
Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
2 - 17	880.4590	1758.9034	1758.9880	-0.0845	1	M.ADLKAFAEQLVNLTVK.E (Ions score 12)
6 - 17	444.8610	1331.5612	1331.7449	-0.1837	0	K.AFAEQLVNLTVK.E (Ions score 42)
6 - 17	666.8020	1331.5894	1331.7449	-0.1554	0	K.AFAEQLVNLTVK.E (Ions score 84)
6 - 17	666.8030	1331.5914	1331.7449	-0.1534	0	K.AFAEQLVNLTVK.E (Ions score 97)
6 - 17	666.8080	1331.6014	1331.7449	-0.1434	0	K.AFAEQLVNLTVK.E (Ions score 75)
6 - 17	666.8130	1331.6114	1331.7449	-0.1334	0	K.AFAEQLVNLTVK.E (Ions score 77)
6 - 17	666.8320	1331.6494	1331.7449	-0.0954	0	K.AFAEQLVNLTVK.E (Ions score 89)
6 - 17	666.8340	1331.6534	1331.7449	-0.0914	0	K.AFAEQLVNLTVK.E (Ions score 88)
6 - 17	666.8400	1331.6654	1331.7449	-0.0794	0	K.AFAEQLVNLTVK.E (Ions score 80)
6 - 17	666.8420	1331.6694	1331.7449	-0.0754	0	K.AFAEQLVNLTVK.E (Ions score 88)
6 - 17	666.8660	1331.7174	1331.7449	-0.0274	0	K.AFAEQLVNLTVK.E (Ions score 94)
6 - 17	666.8780	1331.7414	1331.7449	-0.0034	0	K.AFAEQLVNLTVK.E (Ions score 80)
18 - 27	565.2660	1128.5174	1128.6390	-0.1216	0	K.EVNELATILK.E (Ions score 34)
56 - 63	447.6820	893.3494	893.4858	-0.1364	0	K.SSFDVVLK.S (Ions score 35)
56 - 63	447.6950	893.3754	893.4858	-0.1104	0	K.SSFDVVLK.S (Ions score 34)
56 - 63	447.7160	893.4174	893.4858	-0.0684	0	K.SSFDVVLK.S (Ions score 47)
56 - 63	447.7320	893.4494	893.4858	-0.0364	0	K.SSFDVVLK.S (Ions score 41)
56 - 63	447.7340	893.4534	893.4858	-0.0324	0	K.SSFDVVLK.S (Ions score 36)
56 - 63	447.7350	893.4554	893.4858	-0.0304	0	K.SSFDVVLK.S (Ions score 42)
56 - 63	447.7600	893.5054	893.4858	0.0196	0	K.SSFDVVLK.S (Ions score 32)
86 - 99	730.8450	1459.6754	1460.7181	-1.0426	1	K.EAKDMVDGAPSVVK.E Oxidation (M) (Ions score 36)
89 - 99	567.2240	1132.4334	1132.5434	-0.1100	0	K.DMVDGAPSVVK.E Oxidation (M) (Ions score 41)
113 - 124	644.7600	1287.5054	1287.6558	-0.1503	0	K.TLEEAGAEVELK.- (Ions score 48)
113 - 124	644.7780	1287.5414	1287.6558	-0.1143	0	K.TLEEAGAEVELK.- (Ions score 63)
113 - 124	644.7810	1287.5474	1287.6558	-0.1083	0	K.TLEEAGAEVELK.- (Ions score 51)
113 - 124	644.7840	1287.5534	1287.6558	-0.1023	0	K.TLEEAGAEVELK.- (Ions score 69)
113 - 124	644.7860	1287.5574	1287.6558	-0.0983	0	K.TLEEAGAEVELK.- (Ions score 82)
113 - 124	644.8000	1287.5854	1287.6558	-0.0703	0	K.TLEEAGAEVELK.- (Ions score 68)
113 - 124	644.8140	1287.6134	1287.6558	-0.0423	0	K.TLEEAGAEVELK.- (Ions score 44)
113 - 124	644.8420	1287.6694	1287.6558	0.0137	0	K.TLEEAGAEVELK.- (Ions score 66)



RMS error 154 ppm



RMS error 154 ppm

```
LOCUS       YP_101470                 124 aa            linear   BCT 26-APR-2009
DEFINITION  50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46].
ACCESSION   YP_101470
```

```
VERSION     YP_101470.1  GI:53715478
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 124)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 124)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 124)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from BAD50936.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..124
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..124
                     /product="50S ribosomal protein L7/L12"
                     /calculated_mol_wt=12506
     Region          2..123
                     /region_name="Ribosomal_L7_L12"
                     /note="Ribosomal protein L7/L12. Ribosomal protein L7/L12
                     refers to the large ribosomal subunit proteins L7 and L12,
                     which are identical except that L7 is acetylated at the N
                     terminus. It is a component of the L7/L12 stalk, which is
                     located at the surface of...; cd00387"
                     /db_xref="CDD:100102"
     Site            order(2,15..17,20..21,24,27,42,45..46,48..50,56..57,78,
                     80..81,83,85..86,100..101,103,106)
                     /site_type="other"
                     /note="core dimer interface"
                     /db_xref="CDD:100102"
     Site            order(5,8,12,25,30)
                     /site_type="other"
                     /note="peripheral dimer interface"
                     /db_xref="CDD:100102"
     Site            order(19,22..23,27,30)
                     /site_type="other"
                     /note="L10 interface"
                     /db_xref="CDD:100102"
     Site            order(69..70,73..74,77,84..85,88)
                     /site_type="other"
                     /note="L11 interface"
                     /db_xref="CDD:100102"
     Site            order(70,74,77,84..85,88)
                     /site_type="other"
                     /note="putative EF-Tu interaction site"
                     /db_xref="CDD:100102"
     Site            order(70,73..74,77)
                     /site_type="other"
                     /note="putative EF-G interaction site"
                     /db_xref="CDD:100102"
     CDS             1..124
                     /gene="rplL"
                     /locus_tag="BF4193"
                     /coded_by="complement(NC_006347.1:4790804..4791178)"
                     /note="present in two forms; L12 is normal, while L7 is
                     aminoacylated at the N-terminal serine; the only multicopy
                     ribosomal protein; 4:1 ratio of L7/L12 per ribosome; two
                     L12 dimers bind L10; critically important for translation
                     efficiency and fidelity; stimulates GTPase activity of
                     translation factors"
                     /transl_table=11
                     /db_xref="GeneID:3085275"
```

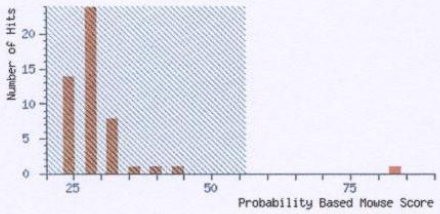
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : LAKSHMY MANICKAN
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 2401
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2401) 10_RK3_01_1421.d\SSP (2401) 10_RK3_01_1421.mgf
Database        : NCBInr 20090522 (8876587 sequences; 3036162093 residues)
Taxonomy        : Bacteria (Eubacteria) (4773688 sequences)
Timestamp       : 30 May 2009 at 10:35:03 GMT
Protein hits    : gi|53713032 GrpE protein [Bacteroides fragilis YCH46]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 56 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53713032 **Mass:** 22007 **Score:** 83 **Queries matched:** 2
   GrpE protein [Bacteroides fragilis YCH46]
   ☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	119	652.2180	1302.4214	1302.6489	-0.2275	0	48	0.3	1	K.SILPVIDDMER.A + Oxidation (M)
☑	240	860.6480	2578.9222	2579.2391	-0.3169	0	35	3.7	1	K.DQPLDTDYHEAIAVIPAPTEEQK.G

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	205	694.9240	2081.7502	2080.9338	0.8164	0	32	8.8	1	TGTGESEEAEEAVGFWQVR
☑	228	1112.9430	2223.8714	2223.0696	0.8019	1	31	11	1	DYLRDDVDQVLIDSPDAFK
☑	224	737.6170	2209.8292	2209.1062	0.7230	2	30	13	1	NSSNIFQIKKFHTNQTMR + Oxidation (M)
☑	192	989.4900	1976.9654	1975.9455	1.0199	2	30	40	1	DMPEGVDTDKVIQRAMR + Oxidation (M)
☑	227	742.2680	2223.7822	2223.0049	0.7773	0	29	17	1	QMPAGYETQIGEGGGMLSGGQR
☑	124	660.2840	1318.5534	1317.6421	0.9114	0	29	34	1	AAGLPVICMETR
☑	54	362.3120	1083.9142	1084.6142	-0.7000	1	28	45	1	AVPIPGYGRR
☑	162	562.7890	1685.3452	1684.7800	0.5651	0	27	30	1	MFTLADDATCIIGTR
☑	129	671.2260	1340.4374	1339.6884	0.7490	0	27	35	1	EALTQFLPEHR
☑	212	700.9180	2099.7322	2099.0932	0.6389	1	26	39	1	QSLDLNIILVTHDKESMK + Oxidation (M)
☑	106	555.3090	1662.9052	1661.8195	1.0857	2	25	1.3e+02	1	DDWIGQAKALMRDK + Oxidation (M)
☑	144	750.8970	2249.6692	2250.0773	-0.4081	0	25	1.2e+02	1	SSGAAVAMVPFVQNLTPQDMR + 2 Oxidation (M)
☑	109	579.1220	1734.3442	1733.8836	0.4606	0	24	1.4e+02	1	LIETLVSENFTDPTR
☑	78	438.3790	1312.1152	1311.6129	0.5023	2	24	1.6e+02	1	YGESPSQSMKKA
☑	173	910.6630	2728.9672	2728.3160	0.6512	2	24	1.2e+02	1	QYSAQSAAMRETGVANSTLMDPKLK + 2 Oxidation (M)
☑	142	705.5310	2113.5712	2113.0209	0.5503	1	23	1.3e+02	1	MTSTTDRTTDTATAVADVLK + Oxidation (M)
☑	189	650.5210	1948.5412	1947.9915	0.5497	2	23	66	1	SSEAAPRGRGSAPGVAPPVW
☑	223	1105.9120	2209.8094	2210.2232	-0.4138	0	23	70	1	GVLPLALLVLSGSLVLAGCDDK
☑	191	656.3610	1966.0612	1967.1316	-1.0704	1	23	1.1e+02	1	LQAAGIGAVRGGALYPVLNK
☑	188	970.8420	1939.6694	1940.1129	-0.4434	1	23	76	1	QKVIKPLVSEQVVQSMK
☑	195	670.0460	2007.1162	2008.0710	-0.9549	2	23	1.1e+02	1	IVRVGGMMVAFGTGRNVTK + Oxidation (M)
☑	241	866.3330	2595.9772	2596.2519	-0.2748	1	23	68	1	APFTAELDASDIEKGTYPHFMLK + Oxidation (M)
☑	219	1069.6480	3205.9222	3205.5251	0.3971	2	22	1.6e+02	1	EAVHETADGRQFLVMHGDEFDSVVRYAK
☑	217	709.5750	2125.7032	2126.1372	-0.4340	0	22	88	1	VAVVGFAPVANQDIGVDNLTK
☑	169	904.6440	2710.9102	2711.3914	-0.4812	0	22	1.7e+02	1	TELSQYAISVLIDMGLSMGLLWQK + Oxidation (M)
☑	235	766.9700	2297.8882	2298.0223	-0.1341	0	22	93	1	YDADGQPAGYATDADGKPVMTR

Mascot Search Results

Protein View

Match to: **gi|53713032** Score: **83**
GrpE protein [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2401) 10_RK3_01_1421.d\SSP (2401) 10_RK3_01_1421.mgf

Nominal mass (M_r): **22007**; Calculated pI value: **4.71**
NCBI BLAST search of gi|53713032 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60681311 from Bacteroides fragilis NCTC 9343
gi|81315621 from Bacteroides fragilis NCTC 9343
gi|81690697 from Bacteroides fragilis
gi|52215897 from Bacteroides fragilis YCH46
gi|60492745 from Bacteroides fragilis NCTC 9343

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **17%**

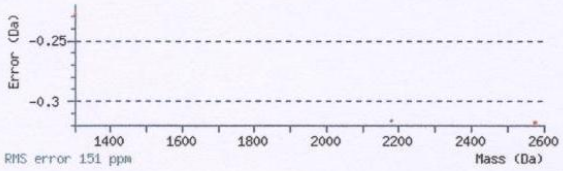
Matched peptides shown in Bold Red

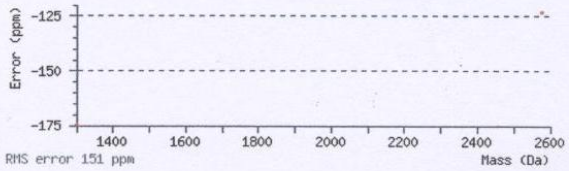
```
  1 MDPKEKKTKQ EEELKVDDIQ DTVEGQSQNE EATEATEPLT AEEKLEKELK
 51 EAQAQIEDQK DKYLRLSAEF DNYRKRTVKE KAELILNGGE KSIKSILPVI
101 DDMERALTTM ETATDVNAVK EGVELIYNKF LSILSQDGVK VIETKDQPLD
151 TDYHEAIAVI PAPTEEQKGK ILDCVQTGYT LNGKVIRHAK VVVGE
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
95 - 105	652.2180	1302.4214	1302.6489	-0.2275	0	K.SILPVIDDMER.A Oxidation (M) (Ions score 48)
146 - 168	860.6480	2578.9222	2579.2391	-0.3169	0	K.DQPLDTDYHEAIAVIPAPTEEQK.G (Ions score 35)






```
LOCUS       YP_099024                195 aa            linear   BCT 26-APR-2009
DEFINITION  GrpE protein [Bacteroides fragilis YCH46].
ACCESSION   YP_099024
VERSION     YP_099024.1  GI:53713032
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 195)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 195)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 195)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD48490.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
```

```
source          1..195
                /organism="Bacteroides fragilis YCH46"
                /strain="YCH46"
                /db_xref="taxon:295405"
Protein         1..195
                /product="GrpE protein"
                /calculated_mol_wt=21832
Region          56..193
                /region_name="GrpE"
                /note="GrpE is the adenine nucleotide exchange factor of
                DnaK (Hsp70)-type ATPases. The GrpE dimer binds to the
                ATPase domain of Hsp70 catalyzing the dissociation of ADP,
                which enables rebinding of ATP, one step in the Hsp70
                reaction cycle in protein folding...; cd00446"
                /db_xref="CDD:73207"
Site            order(58..59,62..63,66,70,73..74,88,92,95,102,106,123,126,
                137)
                /site_type="other"
                /note="dimer interface"
                /db_xref="CDD:73207"
Site            order(75,151,154..158,175,184,188)
                /site_type="other"
                /note="hsp70 (ATPase domain) interactions"
                /db_xref="CDD:73207"
CDS             1..195
                /locus_tag="BF1743"
                /coded_by="complement(NC_006347.1:2050558..2051145)"
                /note="similar to gp:AE016931_34 [Bacteroides
                thetaiotaomicron VPI-5482], percent identity 80 in 195 aa,
                BLASTP E(): 7e-82"
                /transl_table=11
                /db_xref="GeneID:3082995"
```

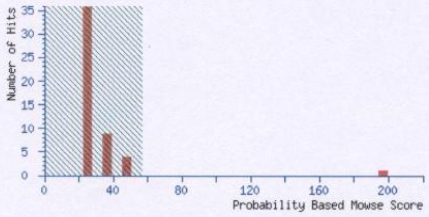
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 2701
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2701) 100_RH7_01_1762.d\SSP (2701) 100_RH7_01_1762.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:17:22 GMT
Protein hits    : gi|53714026 ATP synthase subunit E [Bacteroides fragilis YCH46]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|53714026 **Mass:** 21789 **Score:** 197 **Queries matched:** 4 **emPAI:** 0.54
ATP synthase subunit E [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	120	566.2910	1130.5674	1130.6448	-0.0773	0	58	0.054	1	K.LFAGQAVNALK.S
☑	160	671.2750	1340.5354	1340.6612	* -0.1258	0	40	2.5	1	K.ALFSVSPADGSYK.V
☑	171	740.3800	1478.7454	1478.7955	-0.0501	0	34	12	1	K.AFLRPQLVEMLF.- + Oxidation (M)
☑	192	835.2550	1668.4954	1668.7130	-0.2175	0	65	0.0053	1	K.VNFGEEEFMNYFK.A + Oxidation (M)

Proteins matching the same set of peptides:
gi|255009692 **Mass:** 21804 **Score:** 197 **Queries matched:** 4
putative V-type ATP synthase subunit E [Bacteroides fragilis 3_1_12]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	154	651.8680	1301.7214	1301.6728	0.0487	1	41	2.4	1	VTDLDKAGAWAR
☑	170	738.3230	1474.6314	1475.8209	-1.1894	1	36	6.3	1	TRTTFVQQLINR
☑	132	613.0830	1836.2272	1835.9377	0.2894	0	31	28	1	QSPQPALATIPVSEQDR
☑	187	535.3750	1603.1032	1602.8035	0.2996	1	31	17	1	AEMDAGAVRALDQLK + Oxidation (M)
☑	108	350.2350	1047.6832	1046.5032	1.1799	0	29	37	1	YQSSLLEHA
☑	90	420.6080	1258.8022	1259.6292	-0.8270	2	29	59	1	KEHNKTMQTK + Oxidation (M)
☑	117	551.2270	1650.6592	1650.7777	-0.1185	0	29	73	1	YYASLETFNELSSK
☑	72	378.4320	1132.2742	1132.5989	-0.3247	0	28	53	1	AAGITQAFAQR
☑	246	837.6220	2509.8442	2510.9844	-1.1403	1	28	22	1	ERNFCCGGGSGILMDEMMEIR + 3 Oxidation (M)
☑	163	696.3180	1390.6214	1391.7595	-1.1380	1	28	47	1	FRLNNDLLMLK + Oxidation (M)
☑	212	974.9210	1947.8274	1947.0360	0.7915	1	28	39	1	LLLAFRDGSRPATIMDR + Oxidation (M)
☑	98	468.2630	934.5114	934.4066	0.1049	0	27	69	1	EANCDVVK
☑	131	612.9790	1835.9152	1835.8148	0.1003	1	26	1e+02	1	GFFEAAKDFMGSDATSR
☑	127	599.9810	1796.9212	1796.7934	0.1278	1	26	1.1e+02	1	MHSYSPTGMDSRIGSR + Oxidation (M)
☑	193	843.8680	2528.5822	2529.3663	-0.7841	0	26	1e+02	1	LGQLNSLLAAGQAPPAAGASAAPGTAVR
☑	118	559.4540	1675.3402	1675.8601	-0.5200	2	26	1e+02	1	AASSDRIASDTASRLR
☑	58	351.0500	1050.1282	1049.5036	0.6246	1	26	1.3e+02	1	ARVGSGNSMR + Oxidation (M)
☑	174	752.4730	2254.3972	2254.0504	0.3467	2	25	1.3e+02	1	MSTRDMVVIGMDGRVVEGDR + 2 Oxidation (M)
☑	182	779.3900	2335.1482	2335.1334	0.0148	1	25	1.5e+02	1	MKITELLGAMVQAGMAPSSNDR + Oxidation (M)
☑	56	350.1610	1047.4612	1048.4430	-0.9818	0	25	1.7e+02	1	SDAHMLGMR + 2 Oxidation (M)
☑	215	659.1260	1974.3562	1975.0487	-0.6925	1	25	55	1	DEVVDAAKGLIASGIAPGHR
☑	116	550.4870	1648.4392	1648.8315	-0.3923	2	25	1.3e+02	1	MTNATNARAVARETK + Oxidation (M)
☑	86	414.2450	1239.7132	1238.5457	1.1674	0	25	1.9e+02	1	ISQNMPGMAMK + 2 Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|53714026** Score: **197**
ATP synthase subunit E [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2701) 100_RH7_01_1762.d\SSP (2701) 100_RH7_01_1762.mgf

Nominal mass (M_r): **21789**; Calculated pI value: **5.08**
NCBI BLAST search of gi|53714026 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60682222 from Bacteroides fragilis NCTC 9343
gi|253566982 from Bacteroides sp. 3_2_5
gi|265764376 from Bacteroides sp. 2_1_16
gi|52216891 from Bacteroides fragilis YCH46
gi|60493656 from Bacteroides fragilis NCTC 9343
gi|251944106 from Bacteroides sp. 3_2_5
gi|263256984 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **25%**

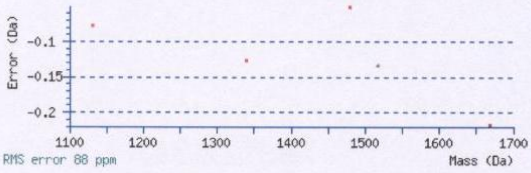
Matched peptides shown in Bold Red

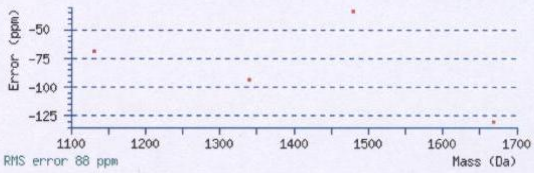
```
  1 MENKIQELTD KIYREGVEKG NEEARRLIAN AQEEAKKIVE DAHKEAESII
 51 ASSRKSADEL TENTKSELKL FAGQAVNALK SEIATMVTDK IVTAPVKEFA
101 QNKDFLNAFI VALASKWSVD EPIIISTSDA ESLKKYFAAN AKALLDKGVT
151 IEQVNGIKAL FSVSPADGSY KVNFGEEEFM NYFKAFLRPQ LVEMLF
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
70 - 80	566.2910	1130.5674	1130.6448	-0.0773	0	K.LFAGQAVNALK.S (Ions score 58)
159 - 171	671.2750	1340.5354	1340.6612	-0.1258	0	K.ALFSVSPADGSYK.V (Ions score 40)
172 - 184	835.2550	1668.4954	1668.7130	-0.2175	0	K.VNFGEEEFMNYFK.A Oxidation (M) (Ions score 65)
185 - 196	740.3800	1478.7454	1478.7955	-0.0501	0	K.AFLRPQLVEMLF.- Oxidation (M) (Ions score 34)






```
LOCUS       YP_100018                196 aa            linear   BCT 26-APR-2009
DEFINITION  ATP synthase subunit E [Bacteroides fragilis YCH46].
ACCESSION   YP_100018
VERSION     YP_100018.1  GI:53714026
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 196)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 196)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 196)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD49484.
            Method: conceptual translation.
```

```
FEATURES             Location/Qualifiers
     source          1..196
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..196
                     /product="ATP synthase subunit E"
                     /calculated_mol_wt=21672
     Region          1..195
                     /region_name="ATP-synt_B"
                     /note="ATP synthase B/B' CF(0); cl07975"
                     /db_xref="CDD:143711"
     Region          18..196
                     /region_name="vATP-synt_E"
                     /note="ATP synthase (E/31 kDa) subunit; pfam01991"
                     /db_xref="CDD:110943"
     CDS             1..196
                     /locus_tag="BF2734"
                     /coded_by="complement(NC_006347.1:3139699..3140289)"
                     /note="similar to gp:AE016931_92 [Bacteroides
                     thetaiotaomicron VPI-5482], percent identity 89 in 196 aa,
                     BLASTP E(): 3e-95"
                     /transl_table=11
                     /db_xref="GeneID:3083533"
```

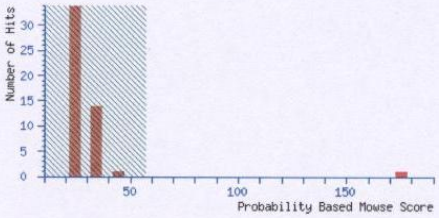
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 2704
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2704) 1000_RG17_01_1193.d\SSP (2704) 1000_RG17_01_1193.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:18:32 GMT
Protein hits    : gi|53714026 ATP synthase subunit E [Bacteroides fragilis YCH46]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|53714026 **Mass:** 21789 **Score:** 175 **Queries matched:** 5 **emPAI:** 0.33
ATP synthase subunit E [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	74	566.3400	1130.6654	1130.6448	0.0207	0	54	0.12	1	K.LFAGQAVNALK.S
☑	75	566.7860	1131.5574	1130.6448*	0.9127	0	(26)	82	1	K.LFAGQAVNALK.S
☑	87	671.3080	1340.6014	1340.6612	-0.0598	0	30	26	1	K.ALFSVSPADGSYK.V
☑	101	740.3580	1478.7014	1478.7955	-0.0941	0	33	12	1	K.AFLRPQLVEMLF.- + Oxidation (M)
☑	155	1059.9960	2117.9774	2117.0892	0.8883	1	57	0.045	1	K.WSVDEPIIISTSDAESLKK.Y

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	100	738.3090	1474.6034	1474.8395	-0.2360	1	49	0.38	1	ELIEISKFLLDR
☑	167	767.3490	2299.0252	2298.2947	0.7305	1	34	7.7	1	QKPQNAQALETLDTIKFLLK
☑	80	651.8460	1301.6774	1301.6463	0.0311	0	31	23	1	VTDIDVNVNEGK
☑	107	786.0440	2355.1102	2354.2376	0.8725	2	31	34	1	TGDVLAAGPVLGKVRAMVDDSGAR
☑	83	657.1320	1312.2494	1312.6987	-0.4492	0	29	23	1	TLPAGTSAGAPATAK
☑	161	1106.0070	2209.9994	2210.1306	-0.1312	1	28	36	1	CQIVYNVLKTHIQFGAYR
☑	35	370.1030	738.1914	737.3014	0.8900	0	28	59	1	SMGGEGGK + Oxidation (M)
☑	163	742.6100	2224.8082	2224.0318	0.7764	2	27	31	1	DKNDFGEIVSEAVEEVCKR
☑	39	379.1450	1134.4132	1135.5696	-1.1564	0	26	1.4e+02	1	HTIYTLMNK + Oxidation (M)
☑	174	1209.6530	3625.9372	3624.9913	0.9459	1	26	89	1	QSNILLTELPLMTPTGTFVINGTERVVVSQIVR
☑	116	836.3120	1670.6094	1671.7351	-1.1257	1	26	52	1	ESFADEHPFYRMK + Oxidation (M)
☑	102	740.9110	2219.7112	2220.1653	-0.4541	0	26	1.1e+02	1	AYVLIFLLGYSLSGMGQNFK
☑	153	694.3570	2080.0492	2079.9248	0.1244	2	25	72	1	MREMMEQAEAQRNAQNK + Oxidation (M)
☑	112	808.6620	2422.9642	2423.3830	-0.4188	0	25	1.1e+02	1	LGQALLGPAGMLSLGGLVLMMILR
☑	108	795.2000	2382.5782	2383.1655	-0.5874	2	24	1.3e+02	1	EDKSIIGYLESKEGNNDFIGR
☑	79	646.9760	1937.9062	1938.9833	-1.0771	1	24	1.7e+02	1	GMIYDSTGKILVANNTSR
☑	71	562.3360	1122.6574	1123.7077	-1.0503	2	24	2.1e+02	1	VVASPAKRALL
☑	136	641.2960	1920.8662	1920.0316	0.8346	0	23	1.1e+02	1	LTLNLNSIDSLNAIYTR
☑	120	842.6610	2524.9612	2524.3618	0.5994	1	23	1.5e+02	1	MPIAPGSTLGILGGGQLGRMTALAAR + Oxidation (M)
☑	124	870.3300	2607.9682	2607.3101	0.6580	0	23	1.6e+02	1	SVLEVSEIINEGLSAFLVNNADMK + Oxidation (M)
☑	88	672.3260	2013.9562	2013.9322	0.0240	0	23	2.7e+02	1	MPAGMSTASPSPPSLPAPMR + 2 Oxidation (M)
☑	144	664.5690	1990.6852	1991.0146	-0.3294	1	23	86	1	KNVPMLNTFDKPELSSR + Oxidation (M)
☑	31	360.2380	1077.6922	1077.6546	0.0375	0	23	3.3e+02	1	LALPPVDVVR
☑	115	827.1690	2478.4852	2478.2074	0.2778	1	23	1.8e+02	1	IRLDGQDLGAQMVQAGQAWSYR + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|53714026** Score: **175**
ATP synthase subunit E [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2704) 1000_RG17_01_1193.d\SSP (2704) 1000_RG17_01_1193.mgf

Nominal mass (M_r): **21789**; Calculated pI value: **5.08**
NCBI BLAST search of gi|53714026 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60682222 from Bacteroides fragilis NCTC 9343
gi|253566982 from Bacteroides sp. 3_2_5
gi|265764376 from Bacteroides sp. 2_1_16
gi|52216891 from Bacteroides fragilis YCH46
gi|60493656 from Bacteroides fragilis NCTC 9343
gi|251944106 from Bacteroides sp. 3_2_5
gi|263256984 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **28%**

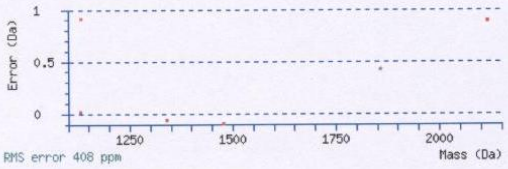
Matched peptides shown in Bold Red

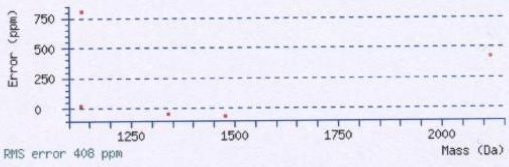
```
  1 MENKIQELTD KIYREGVEKG NEEARRLIAN AQEEAKKIVE DAHKEAESII
 51 ASSRKSADEL TENTKSELKL FAGQAVNALK SEIATMVTDK IVTAPVKEFA
101 QNKDFLNAFI VALASKWSVD EPIIISTSDA ESLKKYFAAN AKALLDKGVT
151 IEQVNGIKAL FSVSPADGSY KVNFGEEEFM NYFKAFLRPQ LVEMLF
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
70 - 80	566.3400	1130.6654	1130.6448	0.0207	0	K.LFAGQAVNALK.S (Ions score 54)
70 - 80	566.7860	1131.5574	1130.6448	0.9127	0	K.LFAGQAVNALK.S (Ions score 26)
117 - 135	1059.9960	2117.9774	2117.0892	0.8883	1	K.WSVDEPIIISTSDAESLKK.Y (Ions score 57)
159 - 171	671.3080	1340.6014	1340.6612	-0.0598	0	K.ALFSVSPADGSYK.V (Ions score 30)
185 - 196	740.3580	1478.7014	1478.7955	-0.0941	0	K.AFLRPQLVEMLF.- Oxidation (M) (Ions score 33)






```
LOCUS       YP_100018                196 aa            linear   BCT 26-APR-2009
DEFINITION  ATP synthase subunit E [Bacteroides fragilis YCH46].
ACCESSION   YP_100018
VERSION     YP_100018.1  GI:53714026
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 196)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 196)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 196)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD49484.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
```

```
source          1..196
                /organism="Bacteroides fragilis YCH46"
                /strain="YCH46"
                /db_xref="taxon:295405"
Protein         1..196
                /product="ATP synthase subunit E"
                /calculated_mol_wt=21672
Region          1..195
                /region_name="ATP-synt_B"
                /note="ATP synthase B/B' CF(0); cl07975"
                /db_xref="CDD:143711"
Region          18..196
                /region_name="vATP-synt_E"
                /note="ATP synthase (E/31 kDa) subunit; pfam01991"
                /db_xref="CDD:110943"
CDS             1..196
                /locus_tag="BF2734"
                /coded_by="complement(NC_006347.1:3139699..3140289)"
                /note="similar to gp:AE016931_92 [Bacteroides
                thetaiotaomicron VPI-5482], percent identity 89 in 196 aa,
                BLASTP E(): 3e-95"
                /transl_table=11
                /db_xref="GeneID:3083533"
```

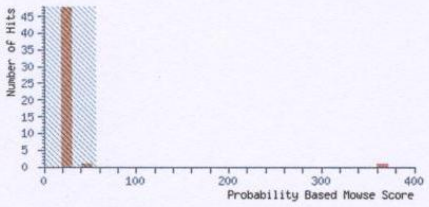
Mascot: http://www.matrixscience.com/

Mascot Search Results

User : LAKSHMY MANICKAN
Email : lakshmy.manickan@unn.ac.uk
Search title : SSP 2804
MS data file : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2804) 10_RJ17_01_1396.d\SSP (2804) 10_RJ17_01_1396.mgf
Database : NCBInr 20090522 (8876587 sequences; 3036162093 residues)
Taxonomy : Bacteria (Eubacteria) (4773688 sequences)
Timestamp : 30 May 2009 at 10:39:02 GMT
Protein hits : gi|53713703 TPR repeat-containing protein [Bacteroides fragilis YCH46]

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 56 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53713703 **Mass:** 46634 **Score:** 366 **Queries matched:** 8 **emPAI:** 0.51
TPR repeat-containing protein [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	125	776.2740	1550.5334	1550.7365	-0.2031	0	64	0.0061	1	K.DNAATWDVAGYIQK.R
☑	140	835.8360	1669.6574	1669.8522	-0.1948	0	74	0.00071	1	K.AEQLINEALTNPETK.D
☑	185	663.5640	1987.6702	1986.9357	0.7345	1	51	0.11	1	K.DKENGQFAMQLLTDAYK.A + Oxidation (M)
☑	186	994.8670	1987.7194	1986.9357	0.7838	1	(26)	81	1	K.DKENGQFAMQLLTDAYK.A + Oxidation (M)
☑	187	995.8420	1989.6694	1989.9030	-0.2335	0	68	0.0022	1	R.VYYNLNMGPEFEEIEK.M + Oxidation (M)
	233	**762.6100**	**2284.8082**	**2283.9738**	**0.8344**	**1**	**19**	**1.7e+02**	**2**	**R.VYYNLNMGPEFEEIEKMM.- + 3 Oxidation (M)**
☑	238	797.0170	2388.0292	2387.2961	0.7331	0	32	9.9	1	K.TILAERPNLINGGIQYFNLNK.N
☑	248	930.6050	2788.7932	2789.2245	-0.4313	0	57	0.025	1	K.FPENQYFFANLVDYYSSSNQNDK.A

Proteins matching the same set of peptides:
gi|60681974 **Mass:** 45773 **Score:** 366 **Queries matched:** 8
hypothetical protein BF2494 [Bacteroides fragilis NCTC 9343]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	221	737.9210	2210.7412	2210.1115	0.6297	0	34	5.2	1	MYILGINSVYHESAVCLLK
☑	131	804.8580	1607.7014	1607.8125	-0.1110	0	33	12	1	MPIYNELMLMPLK + Oxidation (M)
☑	234	766.9800	2297.9182	2299.0790	-1.1609	2	32	9	1	LNGYWEIDKMRSLTDDAEK + Oxidation (M)
☑	80	629.4150	1885.2232	1885.9707	-0.7475	0	32	27	1	SIMDNLGLPVIDDEIVK + Oxidation (M)
☑	121	752.3550	2254.0432	2255.0205	-0.9773	1	30	43	1	LNEPDFDGDDWKIEMFLR + Oxidation (M)
☑	225	1113.0610	2224.1074	2224.0437	0.0638	0	30	21	1	DISWGSTWGAGDNVDIYQLK
☑	44	391.1050	1170.2932	1169.6227	0.6705	0	28	49	1	LLAMEGVNPAR
☑	88	650.7300	1299.4454	1298.7921	0.6533	2	27	37	1	KNESLKLVLQK
☑	157	585.0580	1752.1522	1752.9312	-0.7790	2	27	33	1	VIGGNFKRFQFTPDK
☑	229	747.8320	2240.4742	2240.2277	0.2465	1	26	33	1	YQVDLLTLHATAGRQALTAAK
☑	196	681.0290	2040.0652	2040.9350	-0.8698	2	25	61	1	VEEIPDMYGWKKETAET + Oxidation (M)
☑	145	563.1340	1686.3802	1685.9365	0.4436	2	25	46	1	AGLIHAFFNDRKVAK
☑	204	694.9280	2081.7622	2081.9795	-0.2174	1	25	46	1	MAEMVIGEKVTLEEMGGAR + 2 Oxidation (M)
☑	209	700.9180	2099.7322	2099.0106	0.7216	1	25	45	1	YNIGMAGLIGNDGKFNSNAK + Oxidation (M)
☑	142	838.0410	2511.1012	2511.2329	-0.1317	1	25	1.2e+02	1	SVIYEMHVGGFTRHPNSGVAPEK
☑	213	711.0680	2130.1822	2129.9721	0.2101	2	25	68	1	CDTEEIRKYAICSLSQR
☑	82	632.9100	1895.7082	1896.0693	-0.3612	2	24	1.5e+02	1	FAVPDERAARNIGILVR
☑	108	707.5460	2119.6162	2119.1021	0.5140	2	24	1.2e+02	1	SGEISEQRINEAYQRIVK
☑	129	535.0980	1602.2722	1601.7005	0.5717	1	24	68	1	FDRHAMAGHGESGSK + Oxidation (M)
☑	27	350.2230	1047.6472	1048.5414	-0.8942	0	24	2.1e+02	1	DLNPLANHR
☑	105	703.1500	2106.4282	2105.9299	0.4982	0	24	1.4e+02	1	WDPGAGNGVPGAVMDMDVFR + Oxidation (M)
☑	86	646.1180	1935.3322	1935.0690	0.2632	1	23	1.5e+02	1	APSIESLRQAQPVVVWR

Mascot Search Results

Protein View

Match to: **gi|53713703** Score: **366**
TPR repeat-containing protein [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2804) 10_RJ17_01_1396.d\SSP (2804) 10_RJ17_01_1396.mgf

Nominal mass (M_r): **46634**; Calculated pI value: **5.42**
NCBI BLAST search of gi|53713703 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|253565651 from Bacteroides sp. 3_2_5
gi|265764027 from Bacteroides sp. 2_1_16
gi|52216568 from Bacteroides fragilis YCH46
gi|251945930 from Bacteroides sp. 3_2_5
gi|263256635 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **26%**

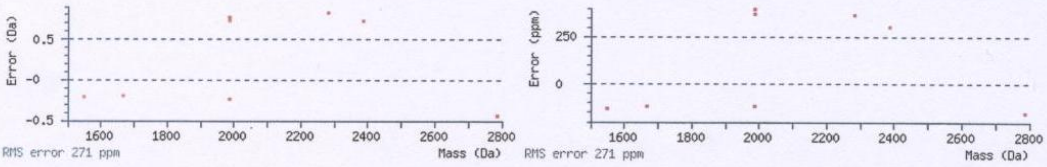
Matched peptides shown in Bold Red

```
  1 MKRVLFSMVL LMAVSFAFAQ EKNVKEAKSI AGEVKPDFAK AEQLINEALT
 51 NPETKDNAAT WDVAGYIQKR INEKEMENAY LRKPYDTLKV YNSVLNMYNY
101 YVKCDELAQI PNEKGKIKNK YRSANSKTIL AERPNLINGG IQYFNLNKNE
151 DALKYFAAYV DAATLPMMEK ENLLEKDTIL PQVAYYATLA ADRVGDKDAV
201 MKYAQYALKD KENGQFAMQL LTDAYKAKGD TAKWVEKLQE GIVKFPENQY
251 FFANLVDYYS SSNQNDKAMQ FADDMLAKDP NNKLYLYVKA YLYHNMKDYE
301 KAIEFYKKTL DIDPAYAEAC SNLGLVYLLQ AQEYADKAPA DINDPNYATA
351 QAEIKKFYEA AKPYYEKARE LKPDQKDLWL QGLYRVYYNL NMGPEFEEIE
401 KMM
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
41 - 55	835.8360	1669.6574	1669.8522	-0.1948	0	K.AEQLINEALTNPETK.D (Ions score 74)
56 - 69	776.2740	1550.5334	1550.7365	-0.2031	0	K.DNAATWDVAGYIQK.R (Ions score 64)
128 - 148	797.0170	2388.0292	2387.2961	0.7331	0	K.TILAERPNLINGGIQYFNLNK.N (Ions score 32)
210 - 226	663.5640	1987.6702	1986.9357	0.7345	1	K.DKENGQFAMQLLTDAYK.A Oxidation (M) (Ions score 51)
210 - 226	994.8670	1987.7194	1986.9357	0.7838	1	K.DKENGQFAMQLLTDAYK.A Oxidation (M) (Ions score 26)
245 - 267	930.6050	2788.7932	2789.2245	-0.4313	0	K.FPENQYFFANLVDYYSSSNQNDK.A (Ions score 57)
386 - 401	995.8420	1989.6694	1989.9030	-0.2335	0	R.VYYNLNMGPEFEEIEK.M Oxidation (M) (Ions score 68)
386 - 403	762.6100	2284.8082	2283.9738	0.8344	1	R.VYYNLNMGPEFEEIEKMM.- 3 Oxidation (M) (Ions score 19)




```
LOCUS       YP_099695                403 aa            linear   BCT 26-APR-2009
DEFINITION  TPR repeat-containing protein [Bacteroides fragilis YCH46].
ACCESSION   YP_099695
VERSION     YP_099695.1  GI:53713703
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 403)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 403)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 403)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
```

```
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD49161.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..403
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..403
                     /product="TPR repeat-containing protein"
                     /calculated_mol_wt=46417
     Region          <180..>380
                     /region_name="PEP_TPR_lipo"
                     /note="putative PEP-CTERM system TPR-repeat lipoprotein;
                     TIGR02917"
                     /db_xref="CDD:131963"
     Region          254..330
                     /region_name="TPR"
                     /note="Tetratricopeptide repeat domain; typically contains
                     34 amino acids
                     [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-
                     X(2)-[ASL]-X(4)-[PKE] is the consensus sequence; found in
                     a variety of organisms including bacteria, cyanobacteria,
                     yeast, fungi...; cd00189"
                     /db_xref="CDD:29151"
     Site            order(254,257..258,287..288,290..291,294..295,321..322,
                     325..326,329..330)
                     /site_type="binding"
                     /note="binding surface"
                     /db_xref="CDD:29151"
     Site            order(256,268,272,275,290,302,306,309,324)
                     /site_type="other"
                     /note="TPR motif"
                     /db_xref="CDD:29151"
     CDS             1..403
                     /locus_tag="BF2412"
                     /coded_by="NC_006347.1:2790670..2791881"
                     /note="similar to gp:AE016929_170 [Bacteroides
                     thetaiotaomicron VPI-5482], percent identity 72 in 397 aa,
                     BLASTP E(): e-171"
                     /transl_table=11
                     /db_xref="GeneID:3083889"
```

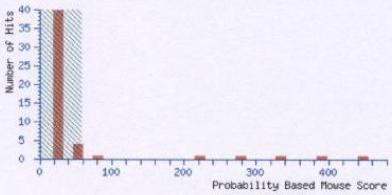
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User          : Lakshmy Manickan
Email         : lakshmy.manickan@unn.ac.uk
Search title  : SSP 2804
MS data file  : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2804) 1000_RE7_01_1712.d\SSP (2804) 1000_RE7_01_1712.mgf
Database      : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy      : Bacteria (Eubacteria) (5690016 sequences)
Timestamp     : 6 Jan 2010 at 13:12:18 GMT
Protein hits  : gi|239636400 dihydrolipoyl dehydrogenase [Staphylococcus warneri L37603]
                gi|53713703  TPR repeat-containing protein [Bacteroides fragilis YCH46]
                gi|223043880 dihydrolipoyl dehydrogenase [Staphylococcus capitis SK14]
                gi|70726857  dihydrolipoamide dehydrogenase [Staphylococcus haemolyticus JCSC1435]
                gi|150011035 dihydrolipoamide dehydrogenase [Staphylococcus xylosus]
                gi|52080062  dihydrolipoamide dehydrogenase [Bacillus licheniformis ATCC 14580]
                gi|50084222  dihydrolipoamide dehydrogenase (E3 component of pyruvate and 2-oxoglutarate dehydrogenase complexes) [Acinetobacte
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|239636400 **Mass:** 49727 **Score:** 447 **Queries matched:** 11 **emPAI:** 0.47
dihydrolipoyl dehydrogenase [Staphylococcus warneri L37603]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	113	493.7310	985.4474	985.5808	-0.1333	0	21	2.9e+02	6	K.LTGGVEGLLK.G
	141	640.7770	1279.5394	1278.6456	0.8939	0	18	4.2e+02	2	R.ALSLDDTNGFVK.L
☑	123	547.5700	1639.6882	1639.8781	-0.1899	0	(22)	2.5e+02	1	R.VIDSTGALNLQEVPGK.L
☑	182	820.8600	1639.7054	1639.8781	-0.1726	0	79	0.00029	1	R.VIDSTGALNLQEVPGK.L
☑	127	561.9150	1682.7232	1682.8839	-0.1607	0	(52)	0.27	1	R.RPNTDELGLEELGLK.F
☑	187	842.3690	1682.7234	1682.8839	-0.1604	0	60	0.026	1	R.RPNTDELGLEELGLK.F
☑	230	1103.5810	2205.1474	2205.2157	-0.0683	0	37	5.3	1	R.TSINNIYAIGDIVPGLPLAHK.A
	169	751.3460	2251.0162	2251.1597	-0.1435	2	22	3.5e+02	4	K.GNKVEIVKGEAYFVDNNSLR.V
☑	238	772.6800	2315.0182	2315.2022	-0.1840	0	104	9.4e-07	1	K.NAIIATGSRPIEIPNFEFGNR.V
☑	246	854.9440	2561.8102	2562.2602	-0.4500	0	43	0.73	1	R.FVEAQHSENLGVIAESVSLNYEK.V
☑	247	863.3940	2587.1602	2586.3694	0.7908	0	68	0.0028	1	M.VVGDFPIETDTIVIGAGPGGYVAAIR.A

2. gi|53713703 **Mass:** 46634 **Score:** 393 **Queries matched:** 7 **emPAI:** 0.41
TPR repeat-containing protein [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	129	582.2570	1162.4994	1162.6135	-0.1140	0	44	1.3	1	K.DLWLQGLYR.V
☑	174	776.2780	1550.5414	1550.7365	-0.1951	0	64	0.0087	1	K.DNAATWDVAGYIQK.R
☑	185	835.8410	1669.6674	1669.8522	-0.1848	0	70	0.0026	1	K.AEQLINEALTNPETK.D
☑	198	893.3540	1784.6934	1784.8443	-0.1509	0	77	0.00044	1	K.VYNSVLNMYNYYVK.C + Oxidation (M)
☑	211	994.4230	1986.8314	1986.9357	-0.1042	1	72	0.0014	1	K.DKENGQFAMQLLTDAYK.A + Oxidation (M)
☑	236	1142.9710	2283.9274	2283.9738	-0.0463	1	46	0.52	1	R.VYYNLNMGPEFEEIEKMM.- + 3 Oxidation (M)
☑	242	797.0000	2387.9782	2387.2961	0.6821	0	21	1.3e+02	1	K.TILAERPNLINGGIQYFNLNK.N

Proteins matching the same set of peptides:
gi|60681974 **Mass:** 45773 **Score:** 393 **Queries matched:** 7
hypothetical protein BF2494 [Bacteroides fragilis NCTC 9343]

3. gi|223043880 **Mass:** 49736 **Score:** 337 **Queries matched:** 10 **emPAI:** 0.38
dihydrolipoyl dehydrogenase [Staphylococcus capitis SK14]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
113	493.7310	985.4474	985.5808	-0.1333	0	21	2.9e+02	6	K.LTGGVEGLLK.G
141	640.7770	1279.5394	1278.6456	0.8939	0	18	4.2e+02	2	R.ALSLDDTNGFVK.L
123	547.5700	1639.6882	1639.8781	-0.1899	0	(22)	2.5e+02	1	R.VIDSTGALNLQEVPGK.L

Mascot Search Results

Protein View

Match to: **gi|60681974** Score: **393**
hypothetical protein BF2494 [Bacteroides fragilis NCTC 9343]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2804) 1000_RE7_01_1712.d\SSP (2804) 1000_RE7_01_1712.mgf

Nominal mass (M_r): **45773**; Calculated pI value: **5.24**
NCBI BLAST search of gi|60681974 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis NCTC 9343
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60493408 from Bacteroides fragilis NCTC 9343

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **27%**

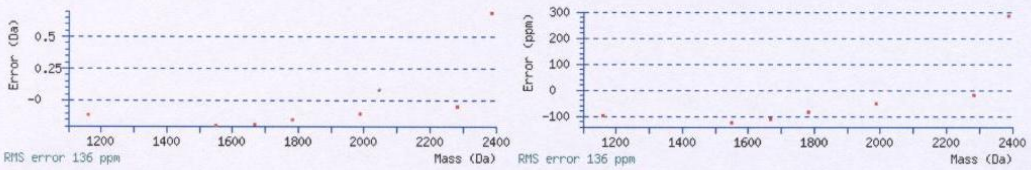
Matched peptides shown in Bold Red

```
  1 MVLLMAVSFA FAQEKNVKEA KSIAGEVKPD FAKAEQLINE ALTNPETKDN
 51 AATWDVAGYI QKRINEKEME NAYLRKPYDT LKVYNSVLNM YNYYVKCDEL
101 AQIPNEKGKI KNKYRSANSK TILAERPNLI NGGIQYFNLN KNEDALKYFA
151 AYVDAATLPM MEKENLLEKD TILPQVAYYA TLAADRVGDK DAVMKYAQYA
201 LKDKENGQFA MQLLTDAYKA KGDTAKWVEK LQEGIVKFPE NQYFFANLVD
251 YYSSSNQNDK AMQFADDMLA KDPNNKLYLY VKAYLYHNMK DYEKAIEFYK
301 KTLDIDPAYA EACSNLGLVY LLQAQEYADK APADINDPNY ATAQAEIKKF
351 YEAAKPYYEK ARELKPDQKD LWLQGLYRVY YNLNMGPEFE EIEKMM
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
34 - 48	835.8410	1669.6674	1669.8522	-0.1848	0	K.AEQLINEALTNPETK.D (Ions score 70)
49 - 62	776.2780	1550.5414	1550.7365	-0.1951	0	K.DNAATWDVAGYIQK.R (Ions score 64)
83 - 96	893.3540	1784.6934	1784.8443	-0.1509	0	K.VYNSVLNMYNYYVK.C Oxidation (M) (Ions score 77)
121 - 141	797.0000	2387.9782	2387.2961	0.6821	0	K.TILAERPNLINGGIQYFNLNK.N (Ions score 21)
203 - 219	994.4230	1986.8314	1986.9357	-0.1042	1	K.DKENGQFAMQLLTDAYK.A Oxidation (M) (Ions score 72)
370 - 378	582.2570	1162.4994	1162.6135	-0.1140	0	K.DLWLQGLYR.V (Ions score 44)
379 - 396	1142.9710	2283.9274	2283.9738	-0.0463	1	R.VYYNLNMGPEFEEIEKMM.- 3 Oxidation (M) (Ions score 46)




```
LOCUS       YP_212118                396 aa            linear   BCT 01-MAY-2009
DEFINITION  hypothetical protein BF2494 [Bacteroides fragilis NCTC 9343].
ACCESSION   YP_212118
VERSION     YP_212118.1  GI:60681974
DBLINK      Project:46
DBSOURCE    REFSEQ: accession NC_003228.3
KEYWORDS    .
SOURCE      Bacteroides fragilis NCTC 9343
  ORGANISM  Bacteroides fragilis NCTC 9343
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 396)
  AUTHORS   Cerdeno-Tarraga,A.M., Patrick,S., Crossman,L.C., Blakely,G.,
            Abratt,V., Lennard,N., Poxton,I., Duerden,B., Harris,B.,
            Quail,M.A., Barron,A., Clark,L., Corton,C., Doggett,J.,
            Holden,M.T., Larke,N., Line,A., Lord,A., Norbertczak,H., Ormond,D.,
            Price,C., Rabbinowitsch,E., Woodward,J., Barrell,B. and Parkhill,J.
  TITLE     Extensive DNA inversions in the B. fragilis genome control variable
            gene expression
  JOURNAL   Science 307 (5714), 1463-1465 (2005)
   PUBMED   15746427
REFERENCE   2  (residues 1 to 396)
  AUTHORS   Cerdeno-Tarraga,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2004) Pathogen Sequencing Unit, Sanger Institute,
            Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United
            Kingdom
REFERENCE   3  (residues 1 to 396)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
```

```
            NCBI review. The reference sequence was derived from CAH08194.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..396
                     /organism="Bacteroides fragilis NCTC 9343"
                     /strain="ATCC 25285; NCTC 9343"
                     /db_xref="ATCC:25285"
                     /db_xref="taxon:272559"
     Protein         1..396
                     /product="hypothetical protein"
                     /calculated_mol_wt=45555
     Region          <173..>373
                     /region_name="PEP_TPR_lipo"
                     /note="putative PEP-CTERM system TPR-repeat lipoprotein;
                     TIGR02917"
                     /db_xref="CDD:131963"
     Region          247..323
                     /region_name="TPR"
                     /note="Tetratricopeptide repeat domain; typically contains
                     34 amino acids
                     [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-
                     X(2)-[ASL]-X(4)-[PKE] is the consensus sequence; found in
                     a variety of organisms including bacteria, cyanobacteria,
                     yeast, fungi...; cd00189"
                     /db_xref="CDD:29151"
     Site            order(247,250..251,280..281,283..284,287..288,314..315,
                     318..319,322..323)
                     /site_type="binding"
                     /note="binding surface"
                     /db_xref="CDD:29151"
     Site            order(249,261,265,268,283,295,299,302,317)
                     /site_type="other"
                     /note="TPR motif"
                     /db_xref="CDD:29151"
     CDS             1..396
                     /locus_tag="BF2494"
                     /coded_by="NC_003228.3:2901913..2903103"
                     /note="Similar to Bacteroides thetaiotaomicron putative
                     tetratricopeptide repeat family protein BT0900
                     SWALL:Q8A9B5 (EMBL:AE016929) (398 aa) fasta scores: E():
                     4.1e-110, 72.79% id in 397 aa"
                     /transl_table=11
                     /db_xref="GeneID:3287481"
```

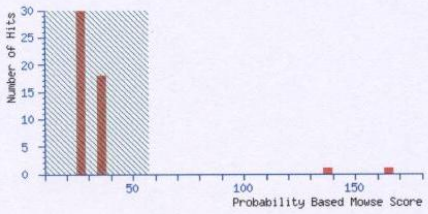
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 2806
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2806) 10_RH14_01_1275.d\SSP (2806) 10_RH14_01_1275.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:19:44 GMT
Protein hits    : gi|29348144  50S ribosomal protein L7/L12 [Bacteroides thetaiotaomicron VPI-5482]
                  gi|189460676 hypothetical protein BACCOP_01323 [Bacteroides coprocola DSM 17136]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|29348144 **Mass:** 12730 **Score:** 165 **Queries matched:** 6 **emPAI:** 0.61
50S ribosomal protein L7/L12 [Bacteroides thetaiotaomicron VPI-5482]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	89	565.2830	1128.5514	1128.6390	-0.0876	0	36	7.8	2	K.EVNELATILK.E
	106	644.2710	1286.5274	1287.6558	-1.1283	0	(14)	1.2e+03	5	K.TLEEAGAEVELK.-
☑	107	644.7870	1287.5594	1287.6558	-0.0963	0	46	0.7	1	K.TLEEAGAEVELK.-
☑	113	666.8140	1331.6134	1331.7449	-0.1314	0	85	0.0001	1	K.AFAEQLVNLTVK.E
☑	114	444.8790	1331.6152	1331.7449	-0.1297	0	(41)	2.5	1	K.AFAEQLVNLTVK.E
☑	115	666.8760	1331.7374	1331.7449	-0.0074	0	(61)	0.024	1	K.AFAEQLVNLTVK.E

Proteins matching the same set of peptides:
gi|53715478 **Mass:** 12688 **Score:** 165 **Queries matched:** 6
50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46]
gi|150003396 **Mass:** 12704 **Score:** 165 **Queries matched:** 6
50S ribosomal protein L7/L12 [Bacteroides vulgatus ATCC 8482]
gi|153805949 **Mass:** 12686 **Score:** 165 **Queries matched:** 6
hypothetical protein BACCAC_00194 [Bacteroides caccae ATCC 43185]
gi|160888407 **Mass:** 12775 **Score:** 165 **Queries matched:** 6
hypothetical protein BACUNI_00820 [Bacteroides uniformis ATCC 8492]
gi|189465410 **Mass:** 12759 **Score:** 165 **Queries matched:** 6
hypothetical protein BACINT_01763 [Bacteroides intestinalis DSM 17393]
gi|198277435 **Mass:** 12757 **Score:** 165 **Queries matched:** 6
hypothetical protein BACPLE_03654 [Bacteroides plebeius DSM 17135]
gi|224023599 **Mass:** 12771 **Score:** 165 **Queries matched:** 6
hypothetical protein BACCOPRO_00306 [Bacteroides coprophilus DSM 18228]
gi|224537808 **Mass:** 12757 **Score:** 165 **Queries matched:** 6
hypothetical protein BACCELL_02695 [Bacteroides cellulosilyticus DSM 14838]

2. gi|189460676 **Mass:** 12771 **Score:** 135 **Queries matched:** 5 **emPAI:** 0.27
hypothetical protein BACCOP_01323 [Bacteroides coprocola DSM 17136]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
89	565.2830	1128.5514	1128.6390	-0.0876	0	36	7.8	2	K.EVNELATILK.E
106	644.2710	1286.5274	1285.6765	0.8509	0	14	1.2e+03	5	K.VLEEAGAEVELK.-
113	666.8140	1331.6134	1331.7449	-0.1314	0	85	0.0001	1	K.AFAEQLVNLTVK.E
114	444.8790	1331.6152	1331.7449	-0.1297	0	(41)	2.5	1	K.AFAEQLVNLTVK.E
115	666.8760	1331.7374	1331.7449	-0.0074	0	(61)	0.024	1	K.AFAEQLVNLTVK.E

Mascot Search Results

Protein View

Match to: **gi|53715478** Score: **165**
50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (2806) 10_RH14_01_1275.d\SSP (2806) 10_RH14_01_1275.mgf

Nominal mass (M_r): **12688**; Calculated pI value: **4.71**
NCBI BLAST search of gi|53715478 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60683451 from Bacteroides fragilis NCTC 9343
gi|253566657 from Bacteroides sp. 3_2_5
gi|255011634 from Bacteroides fragilis 3_1_12
gi|265767535 from Bacteroides sp. 2_1_16
gi|81313501 from Bacteroides fragilis NCTC 9343
gi|81608231 from Bacteroides fragilis
gi|52218343 from Bacteroides fragilis YCH46
gi|60494885 from Bacteroides fragilis NCTC 9343
gi|251944829 from Bacteroides sp. 3_2_5
gi|263252840 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **27%**

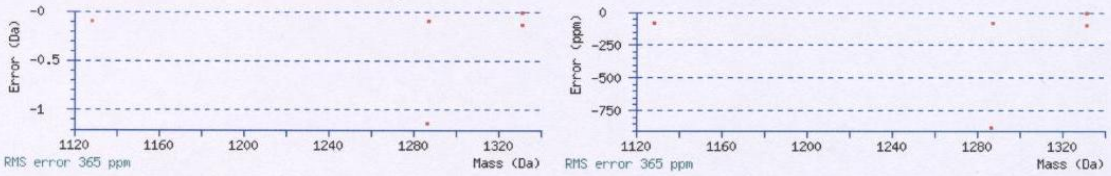
Matched peptides shown in Bold Red

```
  1 MADLKAFAEQ LVNLTVKEVN ELATILKEEY GIEPAAAAVA VAAGPAAGAA
 51 AAEEKSSFDV VLKSAGAAKL QVVKAVKEAC GLGLKEAKDM VDGAPSVVKE
101 GLAKDEAESL KKTLEEAGAE VELK
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence	
6 - 17	666.8140	1331.6134	1331.7449	-0.1314	0	K.AFAEQLVNLTVK.E	(Ions score 85)
6 - 17	444.8790	1331.6152	1331.7449	-0.1297	0	K.AFAEQLVNLTVK.E	(Ions score 41)
6 - 17	666.8760	1331.7374	1331.7449	-0.0074	0	K.AFAEQLVNLTVK.E	(Ions score 61)
18 - 27	565.2830	1128.5514	1128.6390	-0.0876	0	K.EVNELATILK.E	(Ions score 36)
113 - 124	644.2710	1286.5274	1287.6558	-1.1283	0	K.TLEEAGAEVELK.-	(Ions score 14)
113 - 124	644.7870	1287.5594	1287.6558	-0.0963	0	K.TLEEAGAEVELK.-	(Ions score 46)




```
LOCUS       YP_101470                124 aa            linear   BCT 26-APR-2009
DEFINITION  50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46].
ACCESSION   YP_101470
VERSION     YP_101470.1  GI:53715478
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 124)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 124)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 124)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
```

```
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from BAD50936.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..124
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..124
                     /product="50S ribosomal protein L7/L12"
                     /calculated_mol_wt=12506
     Region          2..123
                     /region_name="Ribosomal_L7_L12"
                     /note="Ribosomal protein L7/L12. Ribosomal protein L7/L12
                     refers to the large ribosomal subunit proteins L7 and L12,
                     which are identical except that L7 is acetylated at the N
                     terminus. It is a component of the L7/L12 stalk, which is
                     located at the surface of...; cd00387"
                     /db_xref="CDD:100102"
     Site            order(2,15..17,20..21,24,27,42,45..46,48..50,56..57,78,
                     80..81,83,85..86,100..101,103,106)
                     /site_type="other"
                     /note="core dimer interface"
                     /db_xref="CDD:100102"
     Site            order(5,8,12,25,30)
                     /site_type="other"
                     /note="peripheral dimer interface"
                     /db_xref="CDD:100102"
     Site            order(19,22..23,27,30)
                     /site_type="other"
                     /note="L10 interface"
                     /db_xref="CDD:100102"
     Site            order(69..70,73..74,77,84..85,88)
                     /site_type="other"
                     /note="L11 interface"
                     /db_xref="CDD:100102"
     Site            order(70,74,77,84..85,88)
                     /site_type="other"
                     /note="putative EF-Tu interaction site"
                     /db_xref="CDD:100102"
     Site            order(70,73..74,77)
                     /site_type="other"
                     /note="putative EF-G interaction site"
                     /db_xref="CDD:100102"
     CDS             1..124
                     /gene="rplL"
                     /locus_tag="BF4193"
                     /coded_by="complement(NC_006347.1:4790804..4791178)"
                     /note="present in two forms; L12 is normal, while L7 is
                     aminoacylated at the N-terminal serine; the only multicopy
                     ribosomal protein; 4:1 ratio of L7/L12 per ribosome; two
                     L12 dimers bind L10; critically important for translation
                     efficiency and fidelity; stimulates GTPase activity of
                     translation factors"
                     /transl_table=11
                     /db_xref="GeneID:3085275"
```

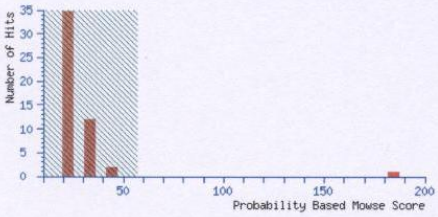
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 3303
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (3303) 10_RJ5_01_1367.d\SSP (3303) 10_RJ5_01_1367.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:24:24 GMT
Protein hits    : gi|53711794 cationic outer membrane protein precursor [Bacteroides fragilis YCH46]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|53711794 **Mass:** 19149 **Score:** 184 **Queries matched:** 8 **emPAI:** 1.24
cationic outer membrane protein precursor [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	110	593.7370	1185.4594	1185.5706	-0.1112	0	20	3.3e+02	7	K.YFGPEGELFK.K
☑	126	657.7750	1313.5354	1313.6656	-0.1301	1	43	1.3	1	R.KYFGPEGELFK.K
	127	438.8580	1313.5522	1313.6656	-0.1134	1	(42)	2.1	2	R.KYFGPEGELFK.K
☑	138	686.3240	1370.6334	1370.7155	-0.0821	0	53	0.14	1	K.FALIDMEYILK.N + Oxidation (M)
	139	686.8900	1371.7654	1370.7155	1.0499	0	(19)	4.1e+02	3	K.FALIDMEYILK.N + Oxidation (M)
☑	197	968.3630	1934.7114	1934.9659	-0.2544	0	35	6.1	1	R.EELMKPIQDEIYNAVK.A + Oxidation (M)
☑	198	645.9190	1934.7352	1934.9659	-0.2307	0	(33)	9.1	1	R.EELMKPIQDEIYNAVK.A + Oxidation (M)
☑	142	697.9610	2090.8612	2091.0670	-0.2058	1	36	12	1	K.REELMKPIQDEIYNAVK.A + Oxidation (M)

Proteins matching the same set of peptides:
gi|255009916 **Mass:** 19218 **Score:** 184 **Queries matched:** 8
putative cationic 19 kDa outer membrane protein precursor [Bacteroides fragilis 3_1_12]
gi|265765175 **Mass:** 19077 **Score:** 184 **Queries matched:** 8
cationic outer membrane protein [Bacteroides sp. 2_1_16]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	127	438.8580	1313.5522	1313.5962	-0.0440	0	43	1.5	1	QYFGPEGELMK + Oxidation (M)
☑	221	700.9150	2099.7232	2100.1262	-0.4031	2	39	2.2	1	YLMVGGDVVDPRVAAQVRR
☑	233	741.9410	2222.8012	2223.1310	-0.3299	2	37	3.2	1	ELLIGKTDYDIFPKEEADK
☑	82	431.2590	1290.7552	1289.6364	1.1188	1	32	40	1	QKPDAEGFDKR
☑	217	694.9340	2081.7802	2080.9371	0.8430	0	31	15	1	SGETIFENNQANDLTGPMK + Oxidation (M)
☑	108	584.9800	1751.9182	1751.6879	0.2302	0	31	37	1	ATCSDACDHLPFSGGR
☑	111	607.0930	1818.2572	1817.9035	0.3537	0	29	47	1	WAATHAPEHWGLMLAK
☑	215	687.9070	2060.6992	2060.0659	0.6333	2	29	22	1	MSRPKKVPMPELPPHER + 2 Oxidation (M)
☑	128	658.3940	1972.1602	1972.9784	-0.8182	0	29	74	1	MVAIGMNSMLLNSYILR + 3 Oxidation (M)
☑	231	1105.9580	2209.9014	2210.2232	-0.3218	0	28	28	1	GVLPLALLVLSGSLVLAGCDDK
☑	235	748.5640	2242.6702	2242.1600	0.5101	2	27	29	1	TIEGRPALPDTRMTTQQRR + Oxidation (M)
☑	109	593.4080	1777.2022	1777.9323	-0.7301	0	27	87	1	TEQVVGTDHPNTLVLR
☑	203	979.4010	2935.1812	2934.4164	0.7648	2	27	66	1	HMGVSALRVINDDMVMPGKGFGTHGHR + Oxidation (M)
☑	134	666.8440	1331.6734	1331.7231	-0.0497	0	27	58	1	AMLQSGSGSLLLR
☑	178	565.9430	1694.8072	1695.8767	-1.0695	0	26	63	1	AKPVAFVSYGGMAGGLR + Oxidation (M)
☑	213	685.9740	2054.9002	2056.0126	-1.1125	1	25	68	1	SGQPHTKAPIYSHVTYDR
☑	230	737.6170	2209.8292	2210.0170	-0.1878	1	25	55	1	MNTKVAIIMGSHSDWETMK + 2 Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|53711794** Score: **184**
cationic outer membrane protein precursor [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (3303) 10_RJ5_01_1367.d\SSP (3303) 10_RJ5_01_1367.mgf

Nominal mass (M_r): **19149**; Calculated pI value: **6.43**
NCBI BLAST search of gi|53711794 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60680025 from Bacteroides fragilis NCTC 9343
gi|253564138 from Bacteroides sp. 3_2_5
gi|52214659 from Bacteroides fragilis YCH46
gi|60491459 from Bacteroides fragilis NCTC 9343
gi|251947914 from Bacteroides sp. 3_2_5

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **22%**

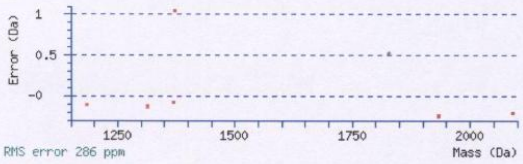
Matched peptides shown in Bold Red

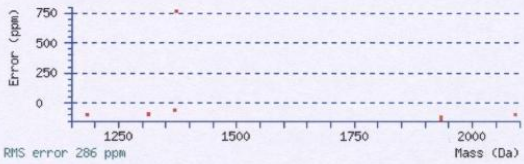
```
  1 MKKSVLFIIL LFAVGMTAQA QKFALIDMEY ILKNIPAYER ANEQLSQATK
 51 QWQGEVEVLA KEAQTMFKDY QAASAKLTAA QKTQKEDAIV EKEKAASELK
101 RKYFGPEGEL FKKREELMKP IQDEIYNAVK AVAEENGYAV VVDRASASSI
151 IFATPRIDVS NEVLAKLGYS N
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
23 - 33	686.3240	1370.6334	1370.7155	-0.0821	0	K.FALIDMEYILK.N Oxidation (M) (Ions score 53)
23 - 33	686.8900	1371.7654	1370.7155	1.0499	0	K.FALIDMEYILK.N Oxidation (M) (Ions score 19)
102 - 112	657.7750	1313.5354	1313.6656	-0.1301	1	R.KYFGPEGELFK.K (Ions score 43)
102 - 112	438.8580	1313.5522	1313.6656	-0.1134	1	R.KYFGPEGELFK.K (Ions score 42)
103 - 112	593.7370	1185.4594	1185.5706	-0.1112	0	K.YFGPEGELFK.K (Ions score 20)
114 - 130	697.9610	2090.8612	2091.0670	-0.2058	1	K.REELMKPIQDEIYNAVK.A Oxidation (M) (Ions score 36)
115 - 130	968.3630	1934.7114	1934.9659	-0.2544	0	R.EELMKPIQDEIYNAVK.A Oxidation (M) (Ions score 35)
115 - 130	645.9190	1934.7352	1934.9659	-0.2307	0	R.EELMKPIQDEIYNAVK.A Oxidation (M) (Ions score 33)






```
LOCUS       YP_097786                171 aa            linear   BCT 26-APR-2009
DEFINITION  cationic outer membrane protein precursor [Bacteroides fragilis
            YCH46].
ACCESSION   YP_097786
VERSION     YP_097786.1  GI:53711794
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 171)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 171)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 171)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD47252.
```

```
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..171
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..171
                     /product="cationic outer membrane protein precursor"
                     /calculated_mol_wt=19030
     Region          4..167
                     /region_name="HlpA"
                     /note="Outer membrane protein [Cell envelope biogenesis,
                     outer membrane]; cl08146"
                     /db_xref="CDD:142087"
     Region          7..168
                     /region_name="OmpH"
                     /note="Outer membrane protein (OmpH-like); pfam03938"
                     /db_xref="CDD:112737"
     CDS             1..171
                     /locus_tag="BF0503"
                     /coded_by="complement(NC_006347.1:578464..578979)"
                     /note="similar to gp:AE016941_210 [Bacteroides
                     thetaiotaomicron VPI-5482], percent identity 71 in 171 aa,
                     BLASTP E(): 2e-63"
                     /transl_table=11
                     /db_xref="GeneID:3082496"
```

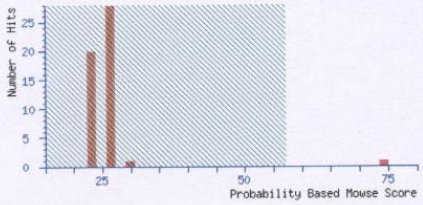
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 3404
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (3404) 10_RJ16_01_1392.d\SSP (3404) 10_RJ16_01_1392.mgf
Database        : NCBInr 20100108 (10291680 sequences; 3511877860 residues)
Taxonomy        : Bacteria (Eubacteria) (5718488 sequences)
Timestamp       : 13 Jan 2010 at 10:33:08 GMT
Protein hits    : gi|143945 Fe-superoxide dismutase [Bacteroides fragilis]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned Decreasing Score Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|143945 **Mass:** 21736 **Score:** 74 **Queries matched:** 2 **emPAI:** 0.15
   Fe-superoxide dismutase [Bacteroides fragilis]
   ☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	138	545.5980	1089.1814	1089.4767	-0.2953	0	14	8.6e+02	7	R.DFGSFENFK.K
☑	174	773.3660	1544.7174	1543.7922	0.9252	0	60	0.048	1	K.LWEIIDWDVVEK.R

Proteins matching the same set of peptides:
gi|533307 **Mass:** 21715 **Score:** 74 **Queries matched:** 2
superoxide dismutase [Bacteroides fragilis]
gi|53713818 **Mass:** 23124 **Score:** 74 **Queries matched:** 2
superoxide dismutase [Bacteroides fragilis YCH46]
gi|60682036 **Mass:** 21758 **Score:** 74 **Queries matched:** 2
superoxide dismutase [Fe] [Bacteroides fragilis NCTC 9343]
gi|150006035 **Mass:** 23299 **Score:** 74 **Queries matched:** 2
superoxide dismutase [Bacteroides vulgatus ATCC 8482]
gi|212692986 **Mass:** 21962 **Score:** 74 **Queries matched:** 2
hypothetical protein BACDOR_02487 [Bacteroides dorei DSM 17855]
gi|237709817 **Mass:** 23300 **Score:** 74 **Queries matched:** 2
superoxide dismutase [Bacteroides sp. 9_1_42FAA]
gi|237725616 **Mass:** 23274 **Score:** 74 **Queries matched:** 2
superoxide dismutase [Bacteroides sp. D4]
gi|255009481 **Mass:** 21757 **Score:** 74 **Queries matched:** 2
superoxide dismutase [Bacteroides fragilis 3_1_12]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	171	509.2150	1524.6232	1523.6596	0.9636	0	28	41	1	MMTQSGNPEIQEK + 2 Oxidation (M)
☑	223	663.3920	1987.1542	1987.9851	-0.8309	1	27	44	1	DAEKAGQNPVDYVVENLK
☑	81	389.8350	1166.4832	1167.5455	-1.0623	0	27	97	1	RPAGTYSGGMR + Oxidation (M)
☑	116	473.3980	944.7814	944.5443	0.2371	1	26	56	1	HKLLGYSK
☑	190	563.4840	1687.4302	1687.9522	-0.5220	1	26	52	1	AIGHIAPIQALKDWR
☑	180	803.5730	2407.6972	2407.3084	0.3888	1	25	1e+02	1	AVAAHVASTAGIAGLNVEGAAFVRR
☑	225	1006.5690	3016.6852	3017.7391	-1.0540	0	25	1.2e+02	1	LLLGIFALFGLATLACAVAPSFISLLVAR
☑	192	847.9400	2540.7982	2540.1826	0.6155	1	25	1.4e+02	1	QGCIESWVSGPGVAADHARTTGQR
☑	204	887.8930	1773.7714	1773.9593	-0.1879	2	24	1.5e+02	1	VRKSQVAMMAEVAQVK
☑	145	590.9250	1769.7532	1769.0352	0.7180	0	24	1.8e+02	1	VIAAVHPGLWQPAVLAK
☑	247	1313.3280	3936.9622	3935.8128	1.1494	2	24	1.1e+02	1	QDHPGQCPKCGMTLEPILPSVDEGENPELIDFRR
☑	240	777.6720	2329.9942	2330.0916	-0.0974	2	24	82	1	SSTQIVEMSTQRMMVKDTDK + Oxidation (M)
☑	183	811.4540	2431.3402	2431.0063	0.3339	1	24	2.1e+02	1	DEYSQYKDYDLMQDFYHR + Oxidation (M)
☑	118	473.9190	1418.7352	1417.7599	0.9753	1	24	2.3e+02	1	AMTIGSKSPLIER + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|60682036** Score: **74**
superoxide dismutase [Fe] [Bacteroides fragilis NCTC 9343]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (3404) 10_RJ16_01_1392.d\SSP (3404) 10_RJ16_01_1392.mgf

Nominal mass (M_r): **21758**; Calculated pI value: **6.07**
NCBI BLAST search of gi|60682036 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis NCTC 9343
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|253565789 from Bacteroides sp. 3_2_5
gi|265764163 from Bacteroides sp. 2_1_16
gi|55977805 from Bacteroides fragilis
gi|60493470 from Bacteroides fragilis NCTC 9343
gi|251946068 from Bacteroides sp. 3_2_5
gi|263256771 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **10%**

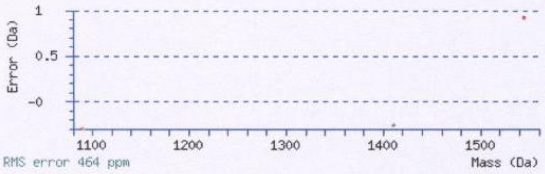
Matched peptides shown in Bold Red

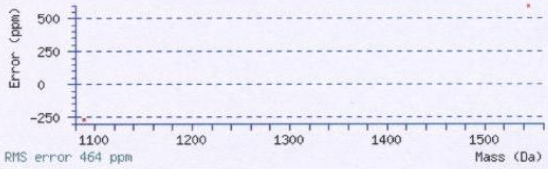
```
  1 MTYEMPKLPY ANNALEPVIS QQTIDYHYGK HLQTYVNNLN SLVPGTEYEG
 51 KTVEAIVASA PDGAIFNNAG QVLNHTLYFL QFAPKPAKNE PAGKLGEAIK
101 RDFGSFENFK KEFNAASVGL FGSGWAWLSV DKDGKLHITK EPNGSNPVRA
151 GLKPLLGFDV WEHAYYLDYQ NRRADHVNKL WEIIDWDVVE KRL
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
102 - 110	545.5980	1089.1814	1089.4767	-0.2953	0	R.DFGSFENFK.K (Ions score 14)
180 - 191	773.3660	1544.7174	1543.7922	0.9252	0	K.LWEIIDWDVVEK.R (Ions score 60)






```
LOCUS       YP_212180                193 aa            linear   BCT 01-MAY-2009
DEFINITION  superoxide dismutase [Fe] [Bacteroides fragilis NCTC 9343].
ACCESSION   YP_212180
VERSION     YP_212180.1  GI:60682036
DBLINK      Project:46
DBSOURCE    REFSEQ: accession NC_003228.3
KEYWORDS    .
SOURCE      Bacteroides fragilis NCTC 9343
  ORGANISM  Bacteroides fragilis NCTC 9343
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 193)
  AUTHORS   Cerdeno-Tarraga,A.M., Patrick,S., Crossman,L.C., Blakely,G.,
            Abratt,V., Lennard,N., Poxton,I., Duerden,B., Harris,B.,
            Quail,M.A., Barron,A., Clark,L., Corton,C., Doggett,J.,
            Holden,M.T., Larke,N., Line,A., Lord,A., Norbertczak,H., Ormond,D.,
            Price,C., Rabbinowitsch,E., Woodward,J., Barrell,B. and Parkhill,J.
  TITLE     Extensive DNA inversions in the B. fragilis genome control variable
            gene expression
  JOURNAL   Science 307 (5714), 1463-1465 (2005)
   PUBMED   15746427
REFERENCE   2  (residues 1 to 193)
  AUTHORS   Cerdeno-Tarraga,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2004) Pathogen Sequencing Unit, Sanger Institute,
            Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United
            Kingdom
REFERENCE   3  (residues 1 to 193)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from CAH08256.
```

```
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..193
                     /organism="Bacteroides fragilis NCTC 9343"
                     /strain="ATCC 25285; NCTC 9343"
                     /db_xref="ATCC:25285"
                     /db_xref="taxon:272559"
     Protein         1..193
                     /product="superoxide dismutase [Fe]"
                     /EC_number="1.15.1.1"
                     /calculated_mol_wt=21640
     Region          1..193
                     /region_name="SodA"
                     /note="Superoxide dismutase [Inorganic ion transport and
                     metabolism]; COG0605"
                     /db_xref="CDD:30950"
     Region          2..83
                     /region_name="Sod_Fe_N"
                     /note="Iron/manganese superoxide dismutases, alpha-hairpin
                     domain; pfam00081"
                     /db_xref="CDD:109149"
     Region          90..192
                     /region_name="Sod_Fe_C"
                     /note="Iron/manganese superoxide dismutases, C-terminal
                     domain; pfam02777"
                     /db_xref="CDD:111650"
     CDS             1..193
                     /gene="sodB"
                     /locus_tag="BF2556"
                     /gene_synonym="sod"
                     /coded_by="NC_003228.3:2982352..2982933"
                     /note="Similar to Bacteroides fragilis superoxide
                     dismutase [Fe] SodB or Sod SWALL:SODF_BACFR (SWALL:P53638)
                     (193 aa) fasta scores: E(): 1.2e-77, 99.48% id in 193 aa,
                     and to Escherichia coli, Escherichia coli O6, Escherichia
                     coli O157:H7, and Shigella flexneri superoxide dismutase
                     [Fe] SodB or B1656 or C2050 or Z2678 or ECS2365 or SF1684
                     or S1816 SWALL:SODF_ECOLI (SWALL:P09157) (192 aa) fasta
                     scores: E(): 1.6e-37, 51.04% id in 192 aa"
                     /transl_table=11
                     /db_xref="GeneID:3287245"
```

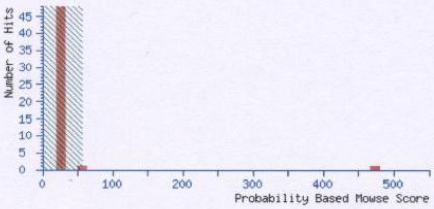
Mascot: http://www.matrixscience.com/

Mascot Search Results

User : Lakshmy Manickan
Email : lakshmy.manickan@unn.ac.uk
Search title : SSP 4202
MS data file : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4202) 10_RN10_01_1584.d\SSP (4202) 10_RN10_01_1584.mgf
Database : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy : Bacteria (Eubacteria) (5690016 sequences)
Timestamp : 6 Jan 2010 at 13:25:46 GMT
Protein hits : gi|53711794 cationic outer membrane protein precursor [Bacteroides fragilis YCH46]

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ Error tolerant

1. gi|53711794 **Mass:** 19149 **Score:** 473 **Queries matched:** 13 **emPAI:** 3.28
cationic outer membrane protein precursor [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	89	593.7220	1185.4294	1185.5706	-0.1412	0	48	0.38	1	K.YFGPEGELFK.K
☑	90	610.7910	1219.5674	1219.6561	-0.0886	0	56	0.071	1	R.ASASSIIFATPR.I
	91	**610.8180**	**1219.6214**	**1219.6561**	**-0.0346**	**0**	**(18)**	**7.2e+02**	**4**	**R.ASASSIIFATPR.I**
☑	92	611.2710	1220.5274	1219.6561	0.8714	0	(45)	0.86	1	R.ASASSIIFATPR.I
☑	105	438.8460	1313.5162	1313.6656	-0.1494	1	49	0.34	1	R.KYFGPEGELFK.K
☑	106	657.7710	1313.5274	1313.6656	-0.1381	1	(42)	1.6	1	R.KYFGPEGELFK.K
☑	113	686.3220	1370.6294	1370.7155	-0.0861	0	73	0.0016	1	K.FALIDMEYILK.N + Oxidation (M)
☑	126	746.7550	1491.4954	1490.7365	0.7589	0	50	0.19	1	K.AVAEENGYAVVVDR.A
☑	167	968.3790	1934.7434	1934.9659	-0.2224	0	37	3.5	1	R.EELMKPIQDEIYNAVK.A + Oxidation (M)
☑	203	697.9660	2090.8762	2091.0670	-0.1908	1	45	0.67	1	K.REELMKPIQDEIYNAVK.A + Oxidation (M)
☑	233	786.3270	2355.9592	2356.2023	-0.2431	1	46	0.44	1	R.ANEQLSQATKQWQGEVEVLAK.E
☑	241	898.4020	2692.1842	2692.3820	-0.1978	1	47	0.36	1	K.AVAEENGYAVVVDRASASSIIFATPR.I
☑	251	991.4350	2971.2832	2971.4385	-0.1554	2	23	85	1	K.QWQGEVEVLAKEAQTMFKDYQAASAK.L + Oxidation (M)

Proteins matching the same set of peptides:
gi|265765175 **Mass:** 19077 **Score:** 473 **Queries matched:** 13
cationic outer membrane protein [Bacteroides sp. 2_1_16]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	100	635.2560	1268.4974	1268.5997	-0.1022	0	40	2.6	1	SVAGDEVADAHAK
☑	182	666.0470	1995.1192	1994.0666	1.0526	0	35	8.3	1	YAYVPYSHFPIGAVLVAK
☑	132	805.8850	1609.7554	1609.9001	-0.1446	0	33	12	1	ISLDTYMSLLVVLK + Oxidation (M)
☑	194	1028.3360	3081.9862	3081.5739	0.4123	1	33	14	1	VACPPRAEIVAGLQAMPPGAEILVAEEHR
☑	234	1181.9720	3542.8942	3542.9397	-0.0456	1	32	20	1	IGLCGLVVSPVGISIVKLFHAPETAIFALAMFR + Oxidation (M)
☑	168	646.8500	1937.5282	1936.8917	0.6364	2	32	11	1	SGADDLLICGPRMMGRR + 2 Oxidation (M)
☑	160	947.3360	2838.9862	2838.3818	0.6044	0	31	26	1	DLANAGGNASDILMNLPSIAVDPEGNVR + Oxidation (M)
☑	175	979.3210	2934.9412	2935.4493	-0.5081	1	29	41	1	QMILMAKMLTSEQDINDLLTHINTF + Oxidation (M)
☑	180	663.0380	1986.0922	1986.0945	-0.0023	1	28	37	1	QARETLSIPVHPMVRPR
☑	173	650.3710	1948.0912	1946.9230	1.1682	0	28	38	1	MIDIHCHIIPGIDDGPK + Oxidation (M)
☑	190	674.2500	2019.7282	2019.8713	-0.1431	1	28	27	1	CPAFAAHAAEGMARAMADR + Oxidation (M)
☑	110	673.9490	1345.8834	1345.7388	0.1447	0	28	81	1	GLTLAQTSVMGIR
☑	108	662.2560	1322.4974	1322.6983	-0.2008	1	28	42	1	VFDGLNKDLFR
☑	246	957.2550	2868.7432	2868.3283	0.4149	1	26	33	1	MADSLSHIAQSGARGFYEGDLAQAMAR + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|53711794** Score: **473**
cationic outer membrane protein precursor [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4202) 10_RN10_01_1584.d\SSP (4202) 10_RN10_01_1584.mgf

Nominal mass (M_r): **19149**; Calculated pI value: **6.43**
NCBI BLAST search of gi|53711794 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60680025 from Bacteroides fragilis NCTC 9343
gi|253564138 from Bacteroides sp. 3_2_5
gi|52214659 from Bacteroides fragilis YCH46
gi|60491459 from Bacteroides fragilis NCTC 9343
gi|251947914 from Bacteroides sp. 3_2_5

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **59%**

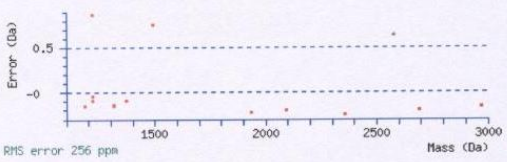
Matched peptides shown in Bold Red

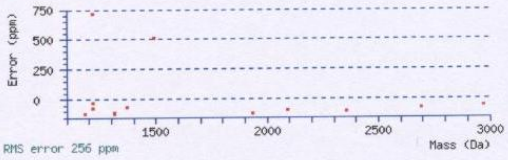
```
  1 MKKSVLFIIL LFAVGMTAQA QKFALIDMEY ILKNIPAYER ANEQLSQATK
 51 QWQGEVEVLA KEAQTMFKDY QAASAKLTAA QKTQKEDAIV EKEKAASELK
101 RKYFGPEGEL FKKREELMKP IQDEIYNAVK AVAEENGYAV VVDRASASSI
151 IFATPRIDVS NEVLAKLGYS N
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
23 - 33	686.3220	1370.6294	1370.7155	-0.0861	0	K.FALIDMEYILK.N Oxidation (M) (Ions score 73)
41 - 61	786.3270	2355.9592	2356.2023	-0.2431	1	R.ANEQLSQATKQWQGEVEVLAK.E (Ions score 46)
51 - 76	991.4350	2971.2832	2971.4385	-0.1554	2	K.QWQGEVEVLAKEAQTMFKDYQAASAK.L Oxidation (M) (Ions score 23)
102 - 112	438.8460	1313.5162	1313.6656	-0.1494	1	R.KYFGPEGELFK.K (Ions score 49)
102 - 112	657.7710	1313.5274	1313.6656	-0.1381	1	R.KYFGPEGELFK.K (Ions score 42)
103 - 112	593.7220	1185.4294	1185.5706	-0.1412	0	K.YFGPEGELFK.K (Ions score 48)
114 - 130	697.9660	2090.8762	2091.0670	-0.1908	1	K.REELMKPIQDEIYNAVK.A Oxidation (M) (Ions score 45)
115 - 130	968.3790	1934.7434	1934.9659	-0.2224	0	R.EELMKPIQDEIYNAVK.A Oxidation (M) (Ions score 37)
131 - 144	746.7550	1491.4954	1490.7365	0.7589	0	K.AVAEENGYAVVVDR.A (Ions score 50)
131 - 156	898.4020	2692.1842	2692.3820	-0.1978	1	K.AVAEENGYAVVVDRASASSIIFATPR.I (Ions score 47)
145 - 156	610.7910	1219.5674	1219.6561	-0.0886	0	R.ASASSIIFATPR.I (Ions score 56)
145 - 156	610.8180	1219.6214	1219.6561	-0.0346	0	R.ASASSIIFATPR.I (Ions score 18)
145 - 156	611.2710	1220.5274	1219.6561	0.8714	0	R.ASASSIIFATPR.I (Ions score 45)






```
LOCUS       YP_097786                171 aa            linear   BCT 26-APR-2009
DEFINITION  cationic outer membrane protein precursor [Bacteroides fragilis
            YCH46].
ACCESSION   YP_097786
VERSION     YP_097786.1  GI:53711794
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 171)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 171)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 171)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD47252.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
```

```
source          1..171
                /organism="Bacteroides fragilis YCH46"
                /strain="YCH46"
                /db_xref="taxon:295405"
Protein         1..171
                /product="cationic outer membrane protein precursor"
                /calculated_mol_wt=19030
Region          4..167
                /region_name="HlpA"
                /note="Outer membrane protein [Cell envelope biogenesis,
                outer membrane]; cl08146"
                /db_xref="CDD:142087"
Region          7..168
                /region_name="OmpH"
                /note="Outer membrane protein (OmpH-like); pfam03938"
                /db_xref="CDD:112737"
CDS             1..171
                /locus_tag="BF0503"
                /coded_by="complement(NC_006347.1:578464..578979)"
                /note="similar to gp:AE016941_210 [Bacteroides
                thetaiotaomicron VPI-5482], percent identity 71 in 171 aa,
                BLASTP E(): 2e-63"
                /transl_table=11
                /db_xref="GeneID:3082496"
```

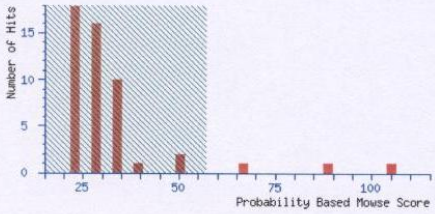
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 4302
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4302) 10_RJ2_01_1361.d\SSP (4302) 10_RJ2_01_1361.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:32:16 GMT
Protein hits    : gi|29349795  adenylate kinase [Bacteroides thetaiotaomicron VPI-5482]
                  gi|260910267 adenylate kinase [Prevotella sp. oral taxon 472 str. F0295]
                  gi|260593579 adenylate kinase [Prevotella veroralis F0319]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|29349795 **Mass:** 21073 **Score:** 105 **Queries matched:** 2
adenylate kinase [Bacteroides thetaiotaomicron VPI-5482]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)*	Delta	Miss	Score	Expect	Rank	Peptide
☑	117	504.2000	1006.3854	1006.5236	-0.1382	0	51	0.23	1	K.GVIFDGFPR.T
☑	171	710.2580	1418.5014	1418.7592	-0.2577	0	54	0.088	1	-.MLNIVIFGAPGSGK.G + Oxidation (M)

Proteins matching the same set of peptides:
gi|53712412 **Mass:** 21018 **Score:** 105 **Queries matched:** 2
adenylate kinase [Bacteroides fragilis YCH46]
gi|150004241 **Mass:** 20871 **Score:** 105 **Queries matched:** 2
adenylate kinase [Bacteroides vulgatus ATCC 8482]
gi|153809272 **Mass:** 21085 **Score:** 105 **Queries matched:** 2
hypothetical protein BACCAC_03584 [Bacteroides caccae ATCC 43185]
gi|154495021 **Mass:** 20930 **Score:** 105 **Queries matched:** 2
hypothetical protein PARMER_04067 [Parabacteroides merdae ATCC 43184]
gi|160887206 **Mass:** 21071 **Score:** 105 **Queries matched:** 2
hypothetical protein BACOVA_05222 [Bacteroides ovatus ATCC 8483]
gi|160891012 **Mass:** 21165 **Score:** 105 **Queries matched:** 2
hypothetical protein BACUNI_03459 [Bacteroides uniformis ATCC 8492]
gi|167765207 **Mass:** 20977 **Score:** 105 **Queries matched:** 2
hypothetical protein BACSTE_03595 [Bacteroides stercoris ATCC 43183]
gi|189462905 **Mass:** 20867 **Score:** 105 **Queries matched:** 2
hypothetical protein BACCOP_03606 [Bacteroides coprocola DSM 17136]
gi|189467969 **Mass:** 21048 **Score:** 105 **Queries matched:** 2
hypothetical protein BACINT_04363 [Bacteroides intestinalis DSM 17393]
gi|198274218 **Mass:** 20879 **Score:** 105 **Queries matched:** 2
hypothetical protein BACPLE_00358 [Bacteroides plebeius DSM 17135]
gi|212690825 **Mass:** 20901 **Score:** 105 **Queries matched:** 2
hypothetical protein BACDOR_00312 [Bacteroides dorei DSM 17855]
gi|218129349 **Mass:** 20933 **Score:** 105 **Queries matched:** 2
hypothetical protein BACEGG_00926 [Bacteroides eggerthii DSM 20697]
gi|218262912 **Mass:** 20944 **Score:** 105 **Queries matched:** 2
hypothetical protein PRABACTJOHN_02899 [Parabacteroides johnsonii DSM 18315]
gi|224025342 **Mass:** 20870 **Score:** 105 **Queries matched:** 2
hypothetical protein BACCOPRO_02081 [Bacteroides coprophilus DSM 18228]
gi|224538342 **Mass:** 21028 **Score:** 105 **Queries matched:** 2

hypothetical protein BACCELL_03233 [Bacteroides cellulosilyticus DSM 14838]
gi|237708208 **Mass:** 21143 **Score:** 105 **Queries matched:** 2
LOW QUALITY PROTEIN: adenylate kinase [Bacteroides sp. 9_1_42FAA]
gi|237715693 **Mass:** 21657 **Score:** 105 **Queries matched:** 2
adenylate kinase [Bacteroides sp. D1]
gi|254884741 **Mass:** 20813 **Score:** 105 **Queries matched:** 2
adenylate kinase [Bacteroides sp. 4_3_47FAA]
gi|255007910 **Mass:** 21018 **Score:** 105 **Queries matched:** 2
adenylate kinase [Bacteroides fragilis 3_1_12]
gi|255693744 **Mass:** 20986 **Score:** 105 **Queries matched:** 2
adenylate kinase [Bacteroides finegoldii DSM 17565]

---

2. gi|260910267 **Mass:** 21197 **Score:** 89 **Queries matched:** 2
adenylate kinase [Prevotella sp. oral taxon 472 str. F0295]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
117	504.2000	1006.3854	1006.5236	-0.1382	0	51	0.23	1	K.GVIFDGFPR.T
171	710.2580	1418.5014	1417.7751	0.7263	1	38	3.7	3	-.MKNIVIFGAPGSGK.G

Proteins matching the same set of peptides:
gi|261880713 **Mass:** 21228 **Score:** 89 **Queries matched:** 2
adenylate kinase [Prevotella bergensis DSM 17361]
gi|281298521 **Mass:** 21183 **Score:** 89 **Queries matched:** 2
adenylate kinase [Prevotella buccalis ATCC 35310]

---

3. gi|260593579 **Mass:** 21280 **Score:** 69 **Queries matched:** 2
adenylate kinase [Prevotella veroralis F0319]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
117	504.2000	1006.3854	1006.5236	-0.1382	0	51	0.23	1	K.GVIFDGFPR.T
105	467.9130	1400.7172	1401.7802	-1.0631	1	18	7.9e+02	5	-.MKNIVIFGAPGAGK.G

Proteins matching the same set of peptides:
gi|252119831 **Mass:** 21127 **Score:** 67 **Queries matched:** 2
adenylate kinase [Prevotella melaninogenica ATCC 25845]
gi|281423652 **Mass:** 20936 **Score:** 67 **Queries matched:** 2
adenylate kinase [Prevotella oris F0302]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	182	496.5760	1486.7062	1485.8151	0.8911	1	35	10	1	QSTTNKQITLPAGK
☑	72	390.1580	778.3014	779.4654	-1.1639	0	33	15	1	IAPAGPVR
☑	222	659.5860	1975.7362	1976.1167	-0.3805	2	31	16	1	LLKIGEEHGAIAGKLTGGGR
☑	224	1002.7190	2003.4234	2004.0316	-0.6082	1	29	33	1	SFIITLGEALHEEFRDK
☑	130	570.2700	1707.7882	1706.8913	0.8969	0	29	72	1	EIMVYSSLIIGVDPR + Oxidation (M)
☑	202	563.1430	1686.4072	1685.7971	0.6101	0	27	39	1	LNEMIFDFNTDTVK
☑	213	949.7190	1897.4234	1896.9074	0.5161	0	26	74	1	MVSIENFGMLSNGQQVK + Oxidation (M)
☑	139	415.1260	1242.3562	1241.6616	0.6946	0	26	57	1	VSIPGSLAGQGEK
☑	189	774.6890	1547.3634	1546.8103	0.5531	1	26	51	1	AGEALSAVGLDGFGKR
☑	220	654.1620	1959.4642	1958.9957	0.4684	1	25	50	1	LMTAYMAFAALESKQIR + Oxidation (M)
☑	128	562.7700	1123.5254	1124.6441	-1.1187	0	25	75	1	EPGVAEIIIGK
☑	113	493.8410	1478.5012	1477.7446	0.7566	2	25	1.7e+02	1	DIMDKETELTRK
☑	211	918.5100	2752.5082	2751.3254	1.1827	2	25	1.5e+02	1	MDAEDAHHLTLRMLGAAGRTGLACAL
☑	216	643.2610	1926.7612	1927.0091	-0.2479	1	24	75	1	TATAAYLKIPDQYTPFK
☑	157	661.2900	1980.8482	1980.1269	0.7213	1	24	2.2e+02	1	SIINKAAAIGGGVIGAGWVAR
☑	124	362.0050	1082.9932	1082.6448	0.3484	1	23	84	1	KVLTEIAGPR
☑	61	358.1450	1071.4132	1072.6063	-1.1931	0	23	3.4e+02	1	QLLLTMAQR
☑	190	776.2270	2325.6592	2325.2403	0.4189	0	23	1.7e+02	1	GSTLPVIFLTGHGDVPMAVSALK + Oxidation (M)
☑	121	354.9720	1061.8942	1062.5128	-0.6186	1	23	1.1e+02	1	ADVQEAMRK + Oxidation (M)
☑	163	689.3310	2064.9712	2065.0184	-0.0472	1	22	3e+02	1	EDKVAQMVSAIQTGGALGMK + 2 Oxidation (M)
☑	136	617.4450	1232.8754	1232.6976	0.1779	1	22	2.4e+02	1	EIKLALASSSSK
☑	125	549.2190	1644.6352	1644.9093	-0.2742	2	22	2.7e+02	1	NLAGSVLVMGKGRATR + Oxidation (M)
☑	215	640.2460	1917.7162	1918.8956	-1.1794	1	22	1.1e+02	1	MTDSPPAANGGREAFVQR + Oxidation (M)
☑	235	1097.0470	2192.0794	2192.1664	-0.0869	1	22	2.2e+02	1	ATNILLAGKTIVVAGYGWCGK
☑	86	422.3290	1263.9652	1263.6360	0.3292	0	22	3.4e+02	1	QQLASAYAAWR
☑	244	820.3240	2457.9502	2458.1984	-0.2482	2	22	1e+02	1	MMYAVGSNERAAELAGAPVKAYK + 2 Oxidation (M)
☑	105	467.9130	1400.7172	1401.7576	-1.0404	1	22	3.5e+02	1	RGETIGIIGGTGSGK
☑	181	743.4140	2227.2202	2227.2259	-0.0057	2	22	3.6e+02	1	AGKNVKAILHQGAHITPTMPK + Oxidation (M)
☑	93	867.4750	866.4677	865.4215	1.0462	0	22	1.7e+02	1	TLMEQTK + Oxidation (M)
☑	219	974.3510	1946.6874	1945.9489	0.7386	0	21	1.3e+02	1	LQGDLDLMPTAVEEMIR + Oxidation (M)
☑	116	497.2930	1488.8572	1489.7810	-0.9239	1	21	4.3e+02	1	TDGVQLVIKMDSGK
☑	184	751.3550	1500.6954	1499.7256	0.9698	0	21	3.8e+02	1	FTSTSVTNFSANPK
☑	241	1162.2730	3483.7972	3482.7472	1.0499	1	21	2.3e+02	1	MNTLPTDNDPVQAWMALGVQSAILARTMPAVR + Oxidation (M)
☑	70	387.1380	772.2614	772.4416	-0.1801	2	21	3.3e+02	1	GASARRR
	242	1168.2350	3501.6832	3501.8752	-0.1921	2	21	2.7e+02	1	VHMASRGHALLADALYGGRPGLGLTRQALHAHR

Mascot Search Results

Protein View

Match to: **gi|53712412** Score: **105**
adenylate kinase [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4302) 10_RJ2_01_1361.d\SSP (4302) 10_RJ2_01_1361.mgf

Nominal mass (M_r): **21018**; Calculated pI value: **5.10**
NCBI BLAST search of gi|53712412 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60680574 from Bacteroides fragilis NCTC 9343
gi|253563556 from Bacteroides sp. 3_2_5
gi|265762597 from Bacteroides sp. 2_1_16
gi|68568759 from Bacteroides fragilis
gi|81316357 from Bacteroides fragilis NCTC 9343
gi|52215277 from Bacteroides fragilis YCH46
gi|60492008 from Bacteroides fragilis NCTC 9343
gi|251947332 from Bacteroides sp. 3_2_5
gi|263255205 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **12%**

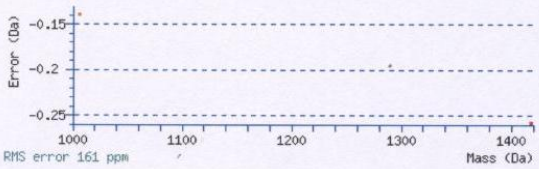
Matched peptides shown in **Bold Red**

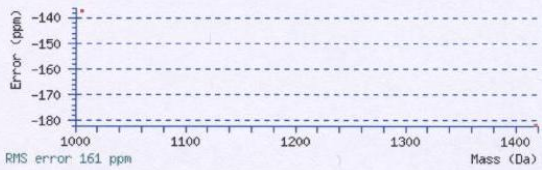
```
  1 MLNIVIFGAP GSGKGTQSER IVEKYGINHI STGDVLRAEI KNGTELGKTA
 51 KGYIDQGQLI PDELMVDILA SVFDSFKDSK GVIFDGFPRT IPQAEALKVM
101 LKERGQDISV MLDLDVPEEE LMTRLIKRGK ESGRADDNEE TIKKRLVVYN
151 TQTSPLKEYY KGEGKYQHIN GLGTMEGIFE DICKAVDTL
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
1 - 14	710.2580	1418.5014	1418.7592	-0.2577	0	-.MLNIVIFGAPGSGK.G Oxidation (M) (Ions score 54)
81 - 89	504.2000	1006.3854	1006.5236	-0.1382	0	K.GVIFDGFPR.T (Ions score 51)






```
LOCUS       YP_098404                189 aa            linear   BCT 26-APR-2009
DEFINITION  adenylate kinase [Bacteroides fragilis YCH46].
ACCESSION   YP_098404
VERSION     YP_098404.1  GI:53712412
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 189)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 189)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 189)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from BAD47870.
            Method: conceptual translation.
```

```
FEATURES             Location/Qualifiers
     source          1..189
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..189
                     /product="adenylate kinase"
                     /EC_number="2.7.4.3"
                     /calculated_mol_wt=20842
     Region          2..187
                     /region_name="adk"
                     /note="adenylate kinase; Reviewed; PRK00279"
                     /db_xref="CDD:134202"
     Region          3..178
                     /region_name="ADK"
                     /note="Adenylate kinase (ADK) catalyzes the reversible
                     phosphoryl transfer from adenosine triphosphates (ATP) to
                     adenosine monophosphates (AMP) and to yield adenosine
                     diphosphates (ADP). This enzyme is required for the
                     biosynthesis of ADP and is essential for...; cd01428"
                     /db_xref="CDD:30189"
     Site            order(32,37,60,85..86,88..89,93)
                     /site_type="other"
                     /note="AMP-binding site"
                     /db_xref="CDD:30189"
     Site            order(37,85,89,128,145,149)
                     /site_type="other"
                     /note="ATP-AMP (Ap5A)-binding site"
                     /db_xref="CDD:30189"
     CDS             1..189
                     /gene="adk"
                     /locus_tag="BF1120"
                     /coded_by="NC_006347.1:1344043..1344612"
                     /note="essential enzyme that recycles AMP in active cells;
                     converts ATP and AMP to two molecules of ADP"
                     /transl_table=11
                     /db_xref="GeneID:3081739"
```

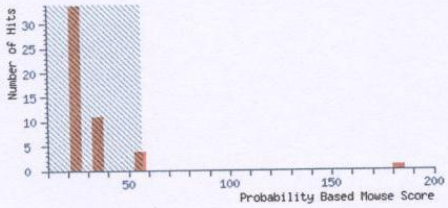
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User          : LAKSHMY MANICKAN
Email         : lakshmy.manickan@unn.ac.uk
Search title  : SSP 4502
MS data file  : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4502) 10_RN2_01_1570.d\SSP (4502) 10_RN2_01_1570.mgf
Database      : NCBInr 20090522 (8876587 sequences; 3036162093 residues)
Taxonomy      : Bacteria (Eubacteria) (4773688 sequences)
Timestamp     : 30 May 2009 at 10:44:47 GMT
Protein hits  : gi|53715240  triosephosphate isomerase [Bacteroides fragilis YCH46]
                gi|21674265  triosephosphate isomerase [Chlorobium tepidum TLS]
                gi|34763378  Triosephosphate isomerase [Fusobacterium nucleatum subsp. vincentii ATCC 49256]
                gi|78223155  triosephosphate isomerase [Geobacter metallireducens GS-15]
                gi|189461507 hypothetical protein BACCOP_02166 [Bacteroides coprocola DSM 17136]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 56 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53715240 **Mass:** 26867 **Score:** 183 **Queries matched:** 4 **emPAI:** 0.42
   triosephosphate isomerase [Bacteroides fragilis YCH46]
   ☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	85	643.2770	1284.5394	1284.6601	-0.1207	0	30	24	1	R.AYYGETVEILK.D
☑	88	655.2560	1308.4974	1308.6714	-0.1740	1	54	0.074	1	K.VADFKGIIDAFN.-
☑	118	808.8830	1615.7514	1615.8974	-0.1459	0	59	0.027	1	K.IVLAYEPVWAIGTGK.T
☑	123	547.2460	1638.7162	1638.8478	-0.1316	0	39	2.7	1	K.TASPAQAQEIHAFIR.S

2. gi|21674265 **Score:** 59 **Queries matched:** 1
   triosephosphate isomerase [Chlorobium tepidum TLS]
   ☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
118	808.8830	1615.7514	1615.8974	-0.1459	0	59	0.027	1	K.LVIAYEPVWAIGTGK.T

Proteins matching the same set of peptides:
gi|218295498 **Score:** 59 **Queries matched:** 1
gi|227987582 **Score:** 59 **Queries matched:** 1
gi|229231135 **Score:** 59 **Queries matched:** 1

3. gi|34763378 **Score:** 59 **Queries matched:** 1
   Triosephosphate isomerase [Fusobacterium nucleatum subsp. vincentii ATCC 49256]
   ☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
118	808.8830	1615.7514	1615.8974	-0.1459	0	59	0.027	1	K.VIIAYEPVWAIGTGK.T

Proteins matching the same set of peptides:
gi|150021144 **Score:** 59 **Queries matched:** 1
gi|197737446 **Score:** 59 **Queries matched:** 1
gi|222055981 **Score:** 59 **Queries matched:** 1
gi|237740886 **Score:** 59 **Queries matched:** 1
gi|237745353 **Score:** 59 **Queries matched:** 1

Mascot Search Results

Protein View

Match to: **gi|53715240** Score: **183**
triosephosphate isomerase [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4502) 10_RN2_01_1570.d\SSP (4502) 10_RN2_01_1570.mgf

Nominal mass (M_r): **26867**; Calculated pI value: **5.14**
NCBI BLAST search of gi|53715240 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60683174 from Bacteroides fragilis NCTC 9343
gi|81313773 from Bacteroides fragilis NCTC 9343
gi|81824915 from Bacteroides fragilis
gi|52218105 from Bacteroides fragilis YCH46
gi|60494608 from Bacteroides fragilis NCTC 9343

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **21%**

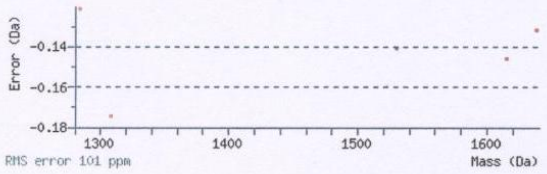
Matched peptides shown in Bold Red

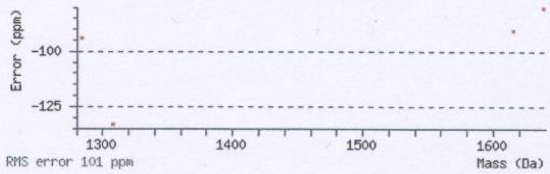
```
  1 MRKNIVAGNW KMNKTLQEGI ALAKELNEAL ANEKPNCDVI ICTPFIHLAS
 51 VTPLVDAAKI GVGAENCADK ESGAYTGEVS AAMVASTGAK YVILGHSERR
101 AYYGETVEIL KDKVKLALAN GLTPIFCIGE VLEEREANKQ NEVVAAQLAS
151 VFDLSAEDFS KIVLAYEPVW AIGTGKTASP AQAQEIHAFI RSAVAEKYGK
201 EIADNTSILY GGSCKPSNAK ELFANPDVDG GLIGGAALKV ADFKGIIDAF
251 N
```

Show predicted peptides also

Sort Peptides By | ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
101 - 111	643.2770	1284.5394	1284.6601	-0.1207	0	R.AYYGETVEILK.D (Ions score 30)
162 - 176	808.8830	1615.7514	1615.8974	-0.1459	0	K.IVLAYEPVWAIGTGK.T (Ions score 59)
177 - 191	547.2460	1638.7162	1638.8478	-0.1316	0	K.TASPAQAQEIHAFIR.S (Ions score 39)
240 - 251	655.2560	1308.4974	1308.6714	-0.1740	1	K.VADFKGIIDAFN.- (Ions score 54)






```
LOCUS       YP_101232                251 aa            linear   BCT 26-APR-2009
DEFINITION  triosephosphate isomerase [Bacteroides fragilis YCH46].
ACCESSION   YP_101232
VERSION     YP_101232.1  GI:53715240
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 251)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 251)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 251)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
```

```
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD50698.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..251
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..251
                     /product="triosephosphate isomerase"
                     /EC_number="5.3.1.1"
                     /calculated_mol_wt=26462
     Region          5..248
                     /region_name="TIM"
                     /note="Triosephosphate isomerase (TIM) is a glycolytic
                     enzyme that catalyzes the interconversion of
                     dihydroxyacetone phosphate and
                     D-glyceraldehyde-3-phosphate. The reaction is very
                     efficient and requires neither cofactors nor metal ions.
                     TIM, usually...; cd00311"
                     /db_xref="CDD:73362"
     Site            order(9,11,96,167,173,213,232,234..235)
                     /site_type="other"
                     /note="substrate binding site"
                     /db_xref="CDD:73362"
     Site            order(9,12,45..47,49,52,65,83,86..87,98..99)
                     /site_type="other"
                     /note="dimer interface"
                     /db_xref="CDD:73362"
     Site            order(11,96,167)
                     /site_type="other"
                     /note="catalytic triad"
                     /db_xref="CDD:73362"
     CDS             1..251
                     /gene="tpiA"
                     /locus_tag="BF3956"
                     /coded_by="complement(NC_006347.1:4511582..4512337)"
                     /note="Reversibly isomerizes the ketone sugar
                     dihydroxyacetone phosphate to the aldehyde sugar
                     glyceraldehyde-3-phosphate"
                     /transl_table=11
                     /db_xref="GeneID:3085037"
```

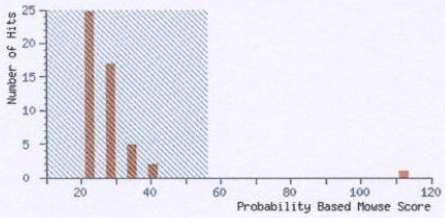
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : LAKSHMY MANICKAN
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 4601
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4601) 10_RJ4_01_1365.d\SSP (4601) 10_RJ4_01_1365.mgf
Database        : NCBInr 20090522 (8876587 sequences; 3036162093 residues)
Taxonomy        : Bacteria (Eubacteria) (4773688 sequences)
Timestamp       : 30 May 2009 at 10:46:21 GMT
Protein hits    : gi|29346739 putative thiol peroxidase [Bacteroides thetaiotaomicron VPI-5482]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 56 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|29346739 **Mass:** 18208 **Score:** 112 **Queries matched:** 2 **emPAI:** 0.19
   putative thiol peroxidase [Bacteroides thetaiotaomicron VPI-5482]
   ☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	104	552.2650	1102.5154	1102.6386	-0.1232	0	46	0.72	1	K.LIGEFIQVGK.V
☑	114	600.7170	1199.4194	1199.6332	*-0.2138	0	67	0.0046	1	R.MADGPLAGLLAR.A + Oxidation (M)

Proteins matching the same set of peptides:
gi|53714055 **Mass:** 18042 **Score:** 112 **Queries matched:** 2
putative thiol peroxidase [Bacteroides fragilis YCH46]
gi|153808277 **Mass:** 18064 **Score:** 112 **Queries matched:** 2
hypothetical protein BACCAC_02565 [Bacteroides caccae ATCC 43185]
gi|160882643 **Mass:** 18079 **Score:** 112 **Queries matched:** 2
hypothetical protein BACOVA_00596 [Bacteroides ovatus ATCC 8483]
gi|212700035 **Mass:** 17981 **Score:** 112 **Queries matched:** 2
hypothetical protein BACFIN_02519 [Bacteroides finegoldii DSM 17565]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	237	766.9440	2297.8102	2298.1380	-0.3278	1	38	2	1	FDFRGVDASIELSGDDTITLK
☑	108	568.4830	1702.4272	1702.8176	-0.3904	1	37	5.8	1	HFGVYEPAQESARGR
☑	225	742.2740	2223.8002	2224.2467	-0.4465	2	34	5.8	1	AKKYLTQVSEGGSLVLFVGTK
☑	195	647.5700	1939.6882	1939.0093	0.6789	1	31	11	1	MLFSMAVGLNAVSMAAKAK
☑	231	748.5940	2242.7602	2243.0892	-0.3290	0	31	10	1	TLDNMGAAPEPNITLLWSER + Oxidation (M)
☑	226	1113.5220	2225.0294	2224.0035	1.0260	2	30	19	1	MCAEVARRGGELVDGDGLMR + 2 Oxidation (M)
☑	241	1171.8740	3512.6002	3511.6639	0.9363	2	27	50	1	NESIRKSIPDLYENNTPIYAECGGLMYLGK + Oxidation (M)
☑	223	1105.9620	2209.9094	2210.2232	-0.3138	0	27	34	1	GVLPLALLVLSGSLVLAGCDDK
☑	230	746.9420	2237.8042	2238.1408	-0.3366	1	27	29	1	MFIGHFAPAFIAAAAYSRGPK + Oxidation (M)
☑	134	679.3990	2035.1752	2035.0347	0.1405	2	25	1.3e+02	1	RYEHRALEAVIQEHER
☑	211	700.9210	2099.7412	2098.9807	0.7605	2	25	47	1	ETNIKETFWESRTDSEK
☑	146	719.7530	2156.2372	2156.1590	0.0782	0	25	1.2e+02	1	VYDLVVIGGGINGVAVAADAAGR
☑	222	737.6270	2209.8592	2211.0221	-1.1629	1	25	51	1	QILCGCGSITEAQADALMKK + Oxidation (M)
☑	89	479.3080	1434.9022	1434.6959	0.2063	0	24	2e+02	1	IQQMEAGMLVQR + 2 Oxidation (M)
☑	111	592.2770	1182.5394	1182.5591	-0.0196	0	24	1e+02	1	MITSVFEDNK
☑	243	1222.4520	2442.8894	2443.3798	-0.4904	2	23	95	1	LGTDAINKIIADPHVLDELKLR
☑	56	366.2100	1095.6082	1096.5149	-0.9067	0	23	2.8e+02	1	IGAGSGSYTER
☑	116	612.3190	1222.6234	1222.6167	0.0068	1	23	1.1e+02	1	LHSTGPERNGR

Mascot Search Results

Protein View

Match to: **gi|53714055** Score: **112**
putative thiol peroxidase [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4601) 10_RJ4_01_1365.d\SSP (4601) 10_RJ4_01_1365.mgf

Nominal mass (M_r): **18042**; Calculated pI value: **5.55**
NCBI BLAST search of gi|53714055 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60682249 from Bacteroides fragilis NCTC 9343
gi|253564697 from Bacteroides sp. 3_2_5
gi|265766326 from Bacteroides sp. 2_1_16
gi|52216920 from Bacteroides fragilis YCH46
gi|60493683 from Bacteroides fragilis NCTC 9343
gi|251946162 from Bacteroides sp. 3_2_5
gi|263253994 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **13%**

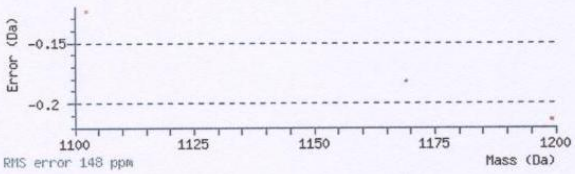
Matched peptides shown in Bold Red

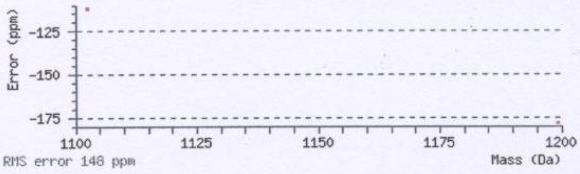
```
  1 MATTNFKGQP VKLIGEFIQV GKVAPDFELV KSDLSSFALK DLKGKNIVLN
 51 IFPSLDTGVC ATSVRKFNKM AAGMKDTVVL AISKDLPFAQ GRFCTTEGIE
101 NVIPLSDFRF SDFDESYGVR MADGPLAGLL ARAVVVIGKD GKVAYTELVP
151 EITQEPDYEK ALAAVK
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
13 - 22	552.2650	1102.5154	1102.6386	-0.1232	0	K.LIGEFIQVGK.V (Ions score 46)
121 - 132	600.7170	1199.4194	1199.6332	-0.2138	0	R.MADGPLAGLLAR.A Oxidation (M) (Ions score 67)






```
LOCUS       YP_100047                166 aa            linear   BCT 26-APR-2009
DEFINITION  putative thiol peroxidase [Bacteroides fragilis YCH46].
ACCESSION   YP_100047
VERSION     YP_100047.1  GI:53714055
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 166)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 166)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 166)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD49513.
```

```
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..166
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..166
                     /product="putative thiol peroxidase"
                     /calculated_mol_wt=17807
     Region          19..164
                     /region_name="PRX_Atyp2cys"
                     /note="Peroxiredoxin (PRX) family, Atypical 2-cys PRX
                     subfamily; composed of PRXs containing peroxidatic and
                     resolving cysteines, similar to the homodimeric thiol
                     specific antioxidant (TSA) protein also known as
                     TRX-dependent thiol peroxidase (Tpx). Tpx is a...;
                     cd03014"
                     /db_xref="CDD:48563"
     Site            order(54,56,85..86,92,108,127..128)
                     /site_type="other"
                     /note="dimer interface"
                     /db_xref="CDD:48563"
     Site            order(57,60,132)
                     /site_type="other"
                     /note="catalytic triad"
                     /db_xref="CDD:48563"
     Site            order(60,94)
                     /site_type="other"
                     /note="peroxidatic and resolving cysteines"
                     /db_xref="CDD:48563"
     CDS             1..166
                     /locus_tag="BF2763"
                     /coded_by="complement(NC_006347.1:3175264..3175764)"
                     /note="similar to gp:AE016931_120 [Bacteroides
                     thetaiotaomicron VPI-5482], percent identity 92 in 166 aa,
                     BLASTP E(): 9e-84"
                     /transl_table=11
                     /db_xref="GeneID:3083562"
```

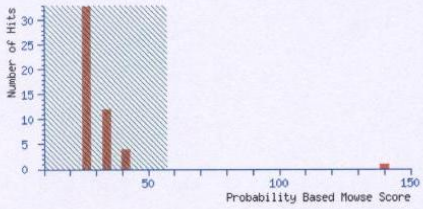
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 4605
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4605) 10_RI7_01_1311.d\SSP (4605) 10_RI7_01_1311.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:36:18 GMT
Protein hits    : gi|53712002 thioredoxin [Bacteroides fragilis YCH46]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53712002 **Mass:** 11637 **Score:** 140 **Queries matched:** 5 **emPAI:** 0.30
   thioredoxin [Bacteroides fragilis YCH46]
   ☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	70	539.7250	1077.4354	1077.6223	-0.1868	0	34	11	1	R.NIPTVLFFK.N
☑	80	601.2670	1200.5194	1200.6424	-0.1229	0	48	0.47	1	K.MVGPIIDELAK.E + Oxidation (M)
☑	81	601.2980	1200.5814	1200.6424	*-0.0609	0	(41)	2.6	1	K.MVGPIIDELAK.E + Oxidation (M)
☑	152	918.8010	1835.5874	1836.7836	-1.1962	0	58	0.031	1	K.CDVDENSDLPAEFGIR.N
	153	612.8830	1835.6272	1836.7836	-1.1564	0	(24)	66	3	K.CDVDENSDLPAEFGIR.N

Proteins matching the same set of peptides:
gi|255007522 **Mass:** 11635 **Score:** 138 **Queries matched:** 5
putative thioredoxin [Bacteroides fragilis 3_1_12]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	90	665.2880	1328.5614	1328.7187	-0.1573	0	38	4.2	1	VINEDTVIADLK
☑	126	819.8490	1637.6834	1637.7897	-0.1062	0	38	3.5	1	SESASFTPNTDIITR
☑	96	692.2880	1382.5614	1382.8245	-0.2631	2	33	13	1	LSIALGKGSKNPAK
☑	83	609.2310	1216.4474	1215.6611	0.7863	1	32	18	1	KAYDPLIGPSR
☑	89	651.7980	1301.5814	1302.7408	-1.1594	1	31	23	1	FIRGIVTDLGGR
☑	227	761.9600	2282.8582	2282.0558	0.8023	2	31	13	1	DMKEASKSIWLDMENQVDK + Oxidation (M)
☑	106	738.2980	1474.5814	1474.8143	-0.2329	1	30	27	1	DAELIKQAFTALR
☑	173	997.4120	1992.8094	1992.8999	-0.0905	2	29	25	1	AGRYKEASEYMPEYQR + Oxidation (M)
☑	153	612.8830	1835.6272	1835.8975	-0.2703	0	28	30	1	CNLSSEPIENAILFTK
☑	71	548.7450	1095.4754	1095.5448	-0.0693	1	28	43	1	DAETLEKYK
☑	87	639.2820	1276.5494	1277.6867	-1.1372	1	27	57	1	GKLIDELFSEK
☑	186	1032.4920	2062.9694	2062.9636	0.0058	2	27	48	1	QRDVQRELAAMSMGDNAR + Oxidation (M)
☑	156	926.7690	1851.5234	1851.8859	-0.3624	1	27	38	1	MKIMTHSNDGFVISEK + Oxidation (M)
☑	166	646.0030	1934.9872	1935.0178	-0.0306	2	26	62	1	KMGGMGTILGLMPGMGKIK + Oxidation (M)
☑	174	665.2790	1992.8152	1991.9478	0.8673	0	26	51	1	MNDLTQTPEMLIATMGAR
☑	240	837.2560	2508.7462	2508.1382	0.6080	2	26	41	1	MAGGAQMFSFPSKSTDMMRIGAR + 2 Oxidation (M)
☑	112	782.0590	2343.1552	2342.1801	0.9751	2	25	1.3e+02	1	REVLAMTAVIYARYSSDNQR
☑	190	1045.0920	3132.2542	3132.4803	-0.2262	0	25	1.2e+02	1	IFEHSGHAPHIEEPEAFMNYYLNFLK
☑	73	551.6520	1651.9342	1651.8286	0.1055	2	25	1.6e+02	1	MNKVIFACVRNAGR + Oxidation (M)
☑	78	572.7520	1143.4894	1143.5448	-0.0554	0	25	1.7e+02	1	VTDDYAVFSK
☑	216	742.2870	2223.8392	2224.2467	-0.4075	2	25	60	1	AKKYLTQVSEGGSLVLFVGTK
☑	213	737.9170	2210.7292	2210.0718	0.6574	0	24	62	1	SEVEFAMYFPEVHLLNAAK + Oxidation (M)
☑	188	696.6030	2086.7872	2087.9768	-1.1897	1	24	73	1	GTDYNMPTQRSMNLGLFK + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|53712002** Score: **140**
thioredoxin [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4605) 10_RI7_01_1311.d\SSP (4605) 10_RI7_01_1311.mgf

Nominal mass (M_r): **11637**; Calculated pI value: **4.51**
NCBI BLAST search of gi|53712002 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60680203 from Bacteroides fragilis NCTC 9343
gi|253563963 from Bacteroides sp. 3_2_5
gi|265765340 from Bacteroides sp. 2_1_16
gi|52214867 from Bacteroides fragilis YCH46
gi|60491637 from Bacteroides fragilis NCTC 9343
gi|251947739 from Bacteroides sp. 3_2_5
gi|263254724 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **34%**

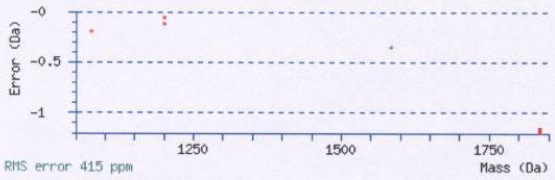
Matched peptides shown in Bold Red

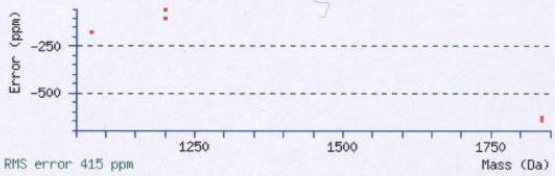
```
  1 MALEITDNNF KEILAEGSPV VIDFWAPWCG PCKMVGPIID ELAKEYEGKV
 51 IMGKCDVDEN SDLPAEFGIR NIPTVLFFKN GELVDKQVGA VGKPAFVEKV
101 EKLL
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence	
34 - 44	601.2670	1200.5194	1200.6424	-0.1229	0	K.MVGPIIDELAK.E Oxidation (M)	(Ions score 48)
34 - 44	601.2980	1200.5814	1200.6424	-0.0609	0	K.MVGPIIDELAK.E Oxidation (M)	(Ions score 41)
55 - 70	918.8010	1835.5874	1836.7836	-1.1962	0	K.CDVDENSDLPAEFGIR.N	(Ions score 58)
55 - 70	612.8830	1835.6272	1836.7836	-1.1564	0	K.CDVDENSDLPAEFGIR.N	(Ions score 24)
71 - 79	539.7250	1077.4354	1077.6223	-0.1868	0	R.NIPTVLFFK.N	(Ions score 34)






```
LOCUS       YP_097994                104 aa            linear   BCT 26-APR-2009
DEFINITION  thioredoxin [Bacteroides fragilis YCH46].
ACCESSION   YP_097994
VERSION     YP_097994.1  GI:53712002
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 104)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 104)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 104)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
```

```
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD47460.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..104
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..104
                     /product="thioredoxin"
                     /calculated_mol_wt=11339
     Region          8..101
                     /region_name="TRX_family"
                     /note="TRX family; composed of two groups: Group I, which
                     includes proteins that exclusively encode a TRX domain;
                     and Group II, which are composed of fusion proteins of TRX
                     and additional domains. Group I TRX is a small ancient
                     protein that alter the redox...; cd02947"
                     /db_xref="CDD:48496"
     Site            order(29,32)
                     /site_type="other"
                     /note="catalytic residues"
                     /db_xref="CDD:48496"
     CDS             1..104
                     /locus_tag="BF0713"
                     /coded_by="NC_006347.1:838887..839201"
                     /note="similar to gp:AE016935_76 [Bacteroides
                     thetaiotaomicron VPI-5482], percent identity 81 in 104 aa,
                     BLASTP E(): 6e-47"
                     /transl_table=11
                     /db_xref="GeneID:3082512"
```

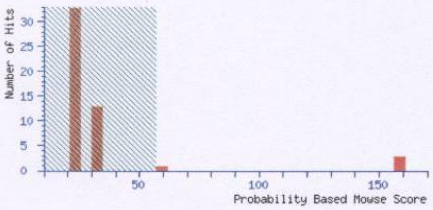
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 4607
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4607) 10_RJ18_01_1401.d\SSP (4607) 10_RJ18_01_1401.mgf
Database        : NCBInr 20100108 (10291680 sequences; 3511877860 residues)
Taxonomy        : Bacteria (Eubacteria) (5718488 sequences)
Timestamp       : 11 Jan 2010 at 10:40:03 GMT
Protein hits    : gi|53715262  malate dehydrogenase [Bacteroides fragilis YCH46]
                  gi|224536272 hypothetical protein BACCELL_01139 [Bacteroides cellulosilyticus DSM 14838]
                  gi|160890687 hypothetical protein BACUNI_03132 [Bacteroides uniformis ATCC 8492]
                  gi|150002856 malate dehydrogenase [Bacteroides vulgatus ATCC 8482]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53715262 **Mass:** 32945 **Score:** 159 **Queries matched:** 4 **emPAI:** 0.21
malate dehydrogenase [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	121	621.2850	1240.5554	1240.7027	-0.1472	0	41	2.3	1	R.EELIGVNAGIVK.S
☑	175	853.8230	1705.6314	1705.8920	-0.2606	0	56	0.055	1	K.LNEVVASTMVGGATLTK.L + Oxidation (M)
	178	862.6630	2584.9672	2585.3121	-0.3449	0	21	2.9e+02	2	K.YSPNAIIVVISNPMDTMTYLALK.S + 2 Oxidation (M)
☑	248	906.0490	2715.1252	2715.4020	-0.2769	0	41	1.2	1	K.LLGTSAWYAPGAAGAYVVESIIHNQK.K

Proteins matching the same set of peptides:
gi|255011805 **Mass:** 32932 **Score:** 159 **Queries matched:** 4
malate dehydrogenase [Bacteroides fragilis 3_1_12]

---

2. gi|224536272 **Mass:** 33016 **Score:** 159 **Queries matched:** 4 **emPAI:** 0.21
hypothetical protein BACCELL_01139 [Bacteroides cellulosilyticus DSM 14838]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
121	621.2850	1240.5554	1240.7027	-0.1472	0	41	2.3	1	R.EELIGVNAGIVK.T
175	853.8230	1705.6314	1705.8920	-0.2606	0	56	0.055	1	K.LNEVVASTMVGGATLTK.L + Oxidation (M)
178	862.6630	2584.9672	2585.3121	-0.3449	0	21	2.9e+02	2	K.YSPNAILVVISNPMDTMTYLALK.S + 2 Oxidation (M)
248	906.0490	2715.1252	2715.4020	-0.2769	0	41	1.2	1	K.LLGTSAWYAPGAAGAYVVESIIHNQK.K

---

3. gi|160890687 **Mass:** 33051 **Score:** 159 **Queries matched:** 4 **emPAI:** 0.21
hypothetical protein BACUNI_03132 [Bacteroides uniformis ATCC 8492]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
121	621.2850	1240.5554	1240.7027	-0.1472	0	41	2.3	1	R.EELIGVNAGIVK.S
175	853.8230	1705.6314	1705.8920	-0.2606	0	56	0.055	1	K.LNEVVASTMVGGATLTK.L + Oxidation (M)
178	862.6630	2584.9672	2585.3121	-0.3449	0	20	3e+02	4	K.YSPNAILVVISNPMDTMTYLSLK.S + Oxidation (M)
248	906.0490	2715.1252	2715.4020	-0.2769	0	41	1.2	1	K.LLGTSAWYAPGAAGAYVVESIIHNQK.K

---

4. gi|150002856 **Mass:** 33174 **Score:** 62 **Queries matched:** 2
malate dehydrogenase [Bacteroides vulgatus ATCC 8482]
☐ Check to include this hit in error tolerant search

Mascot Search Results

Protein View

Match to: **gi|53715262** Score: **159**
malate dehydrogenase [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4607) 10_RJ18_01_1401.d\SSP (4607) 10_RJ18_01_1401.mgf

Nominal mass (M_r): **32945**; Calculated pI value: **5.14**
NCBI BLAST search of gi|53715262 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60683199 from Bacteroides fragilis NCTC 9343
gi|253567153 from Bacteroides sp. 3_2_5
gi|265767749 from Bacteroides sp. 2_1_16
gi|73920985 from Bacteroides fragilis NCTC 9343
gi|73920986 from Bacteroides fragilis
gi|52218127 from Bacteroides fragilis YCH46
gi|60494633 from Bacteroides fragilis NCTC 9343
gi|251943984 from Bacteroides sp. 3_2_5
gi|263252421 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **24%**

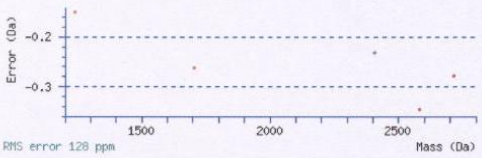
Matched peptides shown in Bold Red

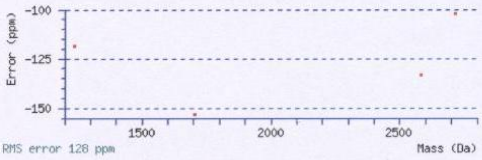
```
  1 MSKVTVVGAG NVGATCANVL AFNEVADEVV MLDVKEGVSE GKAMDMMQTA
 51 QLLGFDTTIV GCTNDYAQTA NSDVVVITSG IPRKPGMTRE ELIGVNAGIV
101 ESVAENLLKY SPNAIIVVIS NPMDTMTYLA LKSLGLPKNR VIGMGGALDS
151 SRFKYFLSQA LGCNANEVEG MVIGGHGDTT MIPLARLATY KGQPVSTLLS
201 EEKLNEVVAS TMVGGATLTK LLGTSAWYAP GAAGAYVVES IIHNQKKMVP
251 CSVMLEGEYG ESDLCIGVPV ILGKNGIEKI VELELNADEK AKFAASAAAV
301 HKTNAALKEV GAL
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
90 - 101	621.2850	1240.5554	1240.7027	-0.1472	0	R.EELIGVNAGIVK.S (Ions score 41)
110 - 132	862.6630	2584.9672	2585.3121	-0.3449	0	K.YSPNAIIVVISNPMDTMTYLALK.S 2 Oxidation (M) (Ions score 21)
204 - 220	853.8230	1705.6314	1705.8920	-0.2606	0	K.LNEVVASTMVGGATLTK.L Oxidation (M) (Ions score 56)
221 - 246	906.0490	2715.1252	2715.4020	-0.2769	0	K.LLGTSAWYAPGAAGAYVVESIIHNQK.K (Ions score 41)






```
LOCUS       YP_101254                313 aa            linear   BCT 26-APR-2009
DEFINITION  malate dehydrogenase [Bacteroides fragilis YCH46].
ACCESSION   YP_101254
VERSION     YP_101254.1  GI:53715262
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 313)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 313)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 313)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD50720.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..313
                     /organism="Bacteroides fragilis YCH46"
```

```
                /strain="YCH46"
                /db_xref="taxon:295405"
Protein         1..313
                /product="malate dehydrogenase"
                /calculated_mol_wt=32545
Region          20..286
                /region_name="LDH-like_MDH"
                /note="L-lactate dehydrogenase-like malate dehydrogenase
                proteins; cd01339"
                /db_xref="CDD:133424"
Site            order(22,37..38,41,44..45,47..48,52,54..56,151..152,154,
                156,221,227..229,232..233,236)
                /site_type="other"
                /note="dimer interface"
                /db_xref="CDD:133424"
Site            order(33..34,78..81,99,119,121,144,148,176)
                /site_type="other"
                /note="NAD(P) binding site"
                /db_xref="CDD:133424"
Site            order(54,56,162,167,169,171,189,191..193,243,245..248,274,
                283)
                /site_type="other"
                /note="tetramer (dimer of dimers) interface"
                /db_xref="CDD:133424"
Site            order(89,121,152,176,215,226)
                /site_type="other"
                /note="substrate binding site"
                /db_xref="CDD:133424"
CDS             1..313
                /locus_tag="BF3978"
                /coded_by="complement(NC_006347.1:4546190..4547131)"
                /note="similar to gp:AE016942_184 [Bacteroides
                thetaiotaomicron VPI-5482], percent identity 92 in 313 aa,
                BLASTP E(): e-163"
                /transl_table=11
                /db_xref="GeneID:3085059"
```

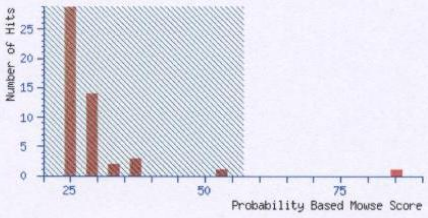
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 4608
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4608) 10_RJ14_01_1388.d\SSP (4608) 10_RJ14_01_1388.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 16:28:30 GMT
Protein hits    : gi|53715262 malate dehydrogenase [Bacteroides fragilis YCH46]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53715262 **Mass:** 32945 **Score:** 85 **Queries matched:** 2 **emPAI:** 0.10
malate dehydrogenase [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	99	621.2790	1240.5434	1240.7027	-0.1592	0	34	12	1	R.EELIGVNAGIVK.S
☑	246	906.0440	2715.1102	2715.4020	-0.2919	0	51	0.12	1	K.LLGTSAWYAPGAAGAYVVESIIHNQK.K

Proteins matching the same set of peptides:
gi|160890687 **Mass:** 33051 **Score:** 85 **Queries matched:** 2
hypothetical protein BACUNI_03132 [Bacteroides uniformis ATCC 8492]
gi|167765305 **Mass:** 33007 **Score:** 85 **Queries matched:** 2
hypothetical protein BACSTE_03693 [Bacteroides stercoris ATCC 43183]
gi|189463781 **Mass:** 32987 **Score:** 85 **Queries matched:** 2
hypothetical protein BACINT_00114 [Bacteroides intestinalis DSM 17393]
gi|224536272 **Mass:** 33016 **Score:** 85 **Queries matched:** 2
hypothetical protein BACCELL_01139 [Bacteroides cellulosilyticus DSM 14838]
gi|255011805 **Mass:** 32932 **Score:** 85 **Queries matched:** 2
malate dehydrogenase [Bacteroides fragilis 3_1_12]
gi|154494237 **Mass:** 33179 **Score:** 83 **Queries matched:** 2
hypothetical protein PARMER_03587 [Parabacteroides merdae ATCC 43184]
gi|218263133 **Mass:** 33190 **Score:** 83 **Queries matched:** 2
hypothetical protein PRABACTJOHN_03033 [Parabacteroides johnsonii DSM 18315]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	245	878.6520	2632.9342	2633.3125	-0.3784	0	53	0.079	1	LPYANNALEPVISQQTIDYHYGK
☑	192	1011.0100	2020.0054	2019.9949	0.0106	1	39	3	1	MGRIFVSFHAAEGGALADR + Oxidation (M)
☑	211	1062.9500	2123.8854	2124.0535	-0.1680	1	38	3.6	1	VPEGLLSPIYGGGGGGHDMRR
☑	136	738.7680	1475.5214	1475.7330	-0.2116	0	35	6.2	1	GGEVFVLDMGEPVK
☑	112	645.4580	1933.3522	1932.1600	1.1921	0	34	16	1	NILFIPILAALIGYFVR
☑	202	694.9150	2081.7232	2080.9405	0.7827	1	33	8.5	1	TGVMQAEDDLNLESMEKR + Oxidation (M)
☑	157	854.1340	2559.3802	2560.2445	-0.8644	1	31	32	1	KVGYESDFFESGGTPSLASSLLNR
☑	108	427.1590	1278.4552	1278.7085	-0.2533	2	30	23	1	RKTGFEGLPFK
☑	121	678.7620	1355.5094	1354.7681	0.7414	2	30	27	1	DRKNVSLTGIPR
☑	137	740.6230	2218.8472	2219.1474	-0.3002	0	28	50	1	DYTLQVVLVDASGTPLYTHK
☑	199	1023.3230	3066.9472	3067.6376	-0.6904	1	28	44	1	LNDIDTLFGLPGIPITDLLRMAQAEGLR + Oxidation (M)
☑	68	468.2760	1401.8062	1400.8286	0.9776	2	28	1.1e+02	1	VRALGLRMALSSK

Mascot Search Results

Protein View

Match to: **gi|53715262** Score: **85**
malate dehydrogenase [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4608) 10_RJ14_01_1388.d\SSP (4608) 10_RJ14_01_1388.mgf

Nominal mass (M_r): **32945**; Calculated pI value: **5.14**
NCBI BLAST search of gi|53715262 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60683199 from Bacteroides fragilis NCTC 9343
gi|253567153 from Bacteroides sp. 3_2_5
gi|265767749 from Bacteroides sp. 2_1_16
gi|73920985 from Bacteroides fragilis NCTC 9343
gi|73920986 from Bacteroides fragilis
gi|52218127 from Bacteroides fragilis YCH46
gi|60494633 from Bacteroides fragilis NCTC 9343
gi|251943984 from Bacteroides sp. 3_2_5
gi|263252421 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **12%**

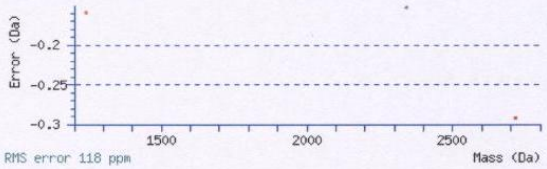
Matched peptides shown in Bold Red

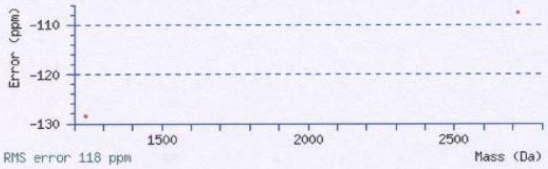
```
  1 MSKVTVVGAG NVGATCANVL AFNEVADEVV MLDVKEGVSE GKAMDMMQTA
 51 QLLGFDTTIV GCTNDYAQTA NSDVVVITSG IPRKPGMTRE ELIGVNAGIV
101 KSVAENLLKY SPNAIIVVIS NPMDTMTYLA LKSLGLPKNR VIGMGGALDS
151 SRFKYFLSQA LGCNANEVEG MVIGGHGDTT MIPLARLATY KGQPVSTLLS
201 EEKLNEVVAS TMVGGATLTK LLGTSAWYAP GAAGAYVVES IIHNQKKMVP
251 CSVMLEGEYG ESDLCIGVPV ILGKNGIEKI VELELNADEK AKFAASAAAV
301 HKTNAALKEV GAL
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
90 - 101	621.2790	1240.5434	1240.7027	-0.1592	0	R.EELIGVNAGIVK.S (Ions score 34)
221 - 246	906.0440	2715.1102	2715.4020	-0.2919	0	K.LLGTSAWYAPGAAGAYVVESIIHNQK.K (Ions score 51)






```
LOCUS       YP_101254                313 aa            linear   BCT 26-APR-2009
DEFINITION  malate dehydrogenase [Bacteroides fragilis YCH46].
ACCESSION   YP_101254
VERSION     YP_101254.1  GI:53715262
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 313)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 313)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 313)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
```

```
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD50720.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..313
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..313
                     /product="malate dehydrogenase"
                     /calculated_mol_wt=32545
     Region          20..286
                     /region_name="LDH-like_MDH"
                     /note="L-lactate dehydrogenase-like malate dehydrogenase
                     proteins; cd01339"
                     /db_xref="CDD:133424"
     Site            order(22,37..38,41,44..45,47..48,52,54..56,151..152,154,
                     156,221,227..229,232..233,236)
                     /site_type="other"
                     /note="dimer interface"
                     /db_xref="CDD:133424"
     Site            order(33..34,78..81,99,119,121,144,148,176)
                     /site_type="other"
                     /note="NAD(P) binding site"
                     /db_xref="CDD:133424"
     Site            order(54,56,162,167,169,171,189,191..193,243,245..248,274,
                     283)
                     /site_type="other"
                     /note="tetramer (dimer of dimers) interface"
                     /db_xref="CDD:133424"
     Site            order(89,121,152,176,215,226)
                     /site_type="other"
                     /note="substrate binding site"
                     /db_xref="CDD:133424"
     CDS             1..313
                     /locus_tag="BF3978"
                     /coded_by="complement(NC_006347.1:4546190..4547131)"
                     /note="similar to gp:AE016942_184 [Bacteroides
                     thetaiotaomicron VPI-5482], percent identity 92 in 313 aa,
                     BLASTP E(): e-163"
                     /transl_table=11
                     /db_xref="GeneID:3085059"
```

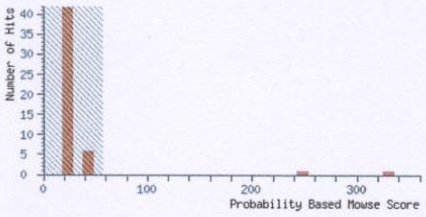
Mascot: http://www.matrixscience.com/

Mascot Search Results

User : Lakshmy Manickan
Email : lakshmy.manickan@unn.ac.uk
Search title : SSP 4702
MS data file : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4702) 10_RJ11_01_1379.d\SSP (4702) 10_RJ11_01_1379.mgf
Database : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy : Bacteria (Eubacteria) (5690016 sequences)
Timestamp : 6 Jan 2010 at 13:37:28 GMT
Protein hits : gi|53715294 elongation factor Ts [Bacteroides fragilis YCH46]
gi|255011774 elongation factor Ts [Bacteroides fragilis 3_1_12]

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ Error tolerant

1. gi|53715294 **Mass:** 36339 **Score:** 331 **Queries matched:** 7 **emPAI:** 0.69
elongation factor Ts [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	122	623.7650	1245.5154	1245.6969	-0.1814	0	68	0.0044	1	K.VEEGFGAIIALK.C
☑	139	686.3340	1370.6534	1370.7405	-0.0871	0	61	0.023	1	K.LTQDILDAAVANK.C
☑	182	847.2930	1692.5714	1693.7539	-1.1824	0	39	2.5	1	K.EVCLLNQEDIMDAK.K + Oxidation (M)
☑	244	772.6100	2314.8082	2315.1679	-0.3597	0	41	1.2	1	K.TLEEVLALPMGDATVAQAVTDR.T + Oxidation (M)
☑	184	849.5650	2545.6732	2546.1239	-0.4508	0	61	0.026	1	K.MELDGYMVLEGATIAAYNHMNR.N + 3 Oxidation (M)
	248	**849.6500**	**2545.9282**	**2546.1239**	**-0.1958**	**0**	**(52)**	**0.089**	**2**	**K.MELDGYMVLEGATIAAYNHMNR.N + 3 Oxidation (M)**
☑	249	854.6580	2560.9522	2561.1989	-0.2467	0	64	0.0067	1	K.QVAMQVAAMNPIAVDEDGVSEEVK.Q + 2 Oxidation (M)

2. gi|255011774 **Mass:** 36340 **Score:** 257 **Queries matched:** 6 **emPAI:** 0.55
elongation factor Ts [Bacteroides fragilis 3_1_12]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	139	686.3340	1370.6534	1370.7405	-0.0871	0	61	0.023	1	K.LTQDILDAAVANK.C
	182	847.2930	1692.5714	1693.7539	-1.1824	0	39	2.5	1	K.EVCLLNQEDIMDAK.K + Oxidation (M)
	244	772.6100	2314.8082	2315.1679	-0.3597	0	41	1.2	1	K.TLEEVLALPMGDATVAQAVTDR.S + Oxidation (M)
	184	849.5650	2545.6732	2546.1239	-0.4508	0	(52)	0.24	2	K.MELDGYMVLEGATISAYNHMNR.N + 2 Oxidation (M)
☑	248	849.6500	2545.9282	2546.1239	-0.1958	0	56	0.04	1	K.MELDGYMVLEGATISAYNHMNR.N + 2 Oxidation (M)
	249	854.6580	2560.9522	2561.1989	-0.2467	0	64	0.0067	1	K.QVAMQVAAMNPIAVDEDGVSEEVK.Q + 2 Oxidation (M)

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	152	742.8350	2225.4832	2225.2168	0.2663	1	34	19	1	RLDVRPLASGVVILTYGPQDG
☑	234	742.3310	2223.9712	2225.0825	-1.1113	2	30	22	1	SSPFGNDTKVGSSAFQQANKR
☑	101	537.8890	1610.6452	1610.8338	-0.1886	1	30	56	1	DKVSQIITYGSMAAK
☑	168	822.2630	1642.5114	1641.7742	0.7372	0	29	25	1	YVELMTSPVIMDGR + 2 Oxidation (M)
☑	64	371.0920	1110.2542	1109.5023	0.7519	0	29	43	1	SMTGDVANATK + Oxidation (M)
☑	132	665.2920	1328.5694	1328.7122	-0.1428	0	27	62	1	QVDIPAMLVATR + Oxidation (M)
☑	233	742.2680	2223.7822	2223.2725	0.5096	2	26	37	1	KNILPQEKLTLLEIEVSEK
☑	138	679.3950	2035.1632	2033.9952	1.1679	1	26	1.2e+02	1	ARNVTIAEFPMSADAAGAAR + Oxidation (M)
☑	208	970.7680	1939.5214	1939.9053	-0.3839	1	25	49	1	MDKEMVIDLVTEQMAR + 2 Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|53715294** Score: **331**
elongation factor Ts [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4702) 10_RJ11_01_1379.d\SSP (4702) 10_RJ11_01_1379.mgf

Nominal mass (M_r): **36339**; Calculated pI value: **5.09**
NCBI BLAST search of gi|53715294 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60683229 from Bacteroides fragilis NCTC 9343
gi|253567185 from Bacteroides sp. 3_2_5
gi|265767781 from Bacteroides sp. 2_1_16
gi|60389524 from Bacteroides fragilis
gi|81313719 from Bacteroides fragilis NCTC 9343
gi|52218159 from Bacteroides fragilis YCH46
gi|60494663 from Bacteroides fragilis NCTC 9343
gi|251944016 from Bacteroides sp. 3_2_5
gi|263252453 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **32%**

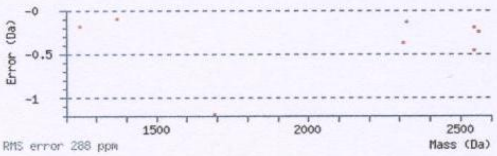
Matched peptides shown in Bold Red

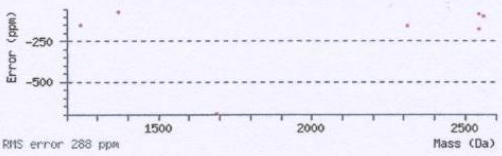
```
  1 MAVTMADITK LRKMTGAGMM DCKNALTDAE GDFDKAMKII REKGQAVAAK
 51 RSDREASEGC VLVKVEEGFG AIIALKCETD FVAQNADFVK LTQDILDAAV
101 ANKCKTLEEV LALPMGDATV AQAVTDRTGI TGEKMELDGY MVLEGATIAA
151 YNHMNRNGLC TMVAFNKKVD EQLAKQVAMQ VAAMNPIAVD EDGVSEEVKQ
201 KEIEVAVEKT KVEQVQKAVE AALKKANINP AHVDSEDHME SNMAKGWITA
251 EDVAKAKEII ATVSAEKAAN MPEQMIQNIA KGRLAKFLKE VCLLNQEDIM
301 DAKKTVREVL KEADPELKVV DFKRFTLRAE
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
65 - 76	623.7650	1245.5154	1245.6969	-0.1814	0	K.VEEGFGAIIALK.C (Ions score 68)
91 - 103	686.3340	1370.6534	1370.7405	-0.0871	0	K.LTQDILDAAVANK.C (Ions score 61)
106 - 127	772.6100	2314.8082	2315.1679	-0.3597	0	K.TLEEVLALPMGDATVAQAVTDR.T Oxidation (M) (Ions score 41)
135 - 156	849.5650	2545.6732	2546.1239	-0.4508	0	K.MELDGYMVLEGATIAAYNHMNR.N 3 Oxidation (M) (Ions score 61)
135 - 156	849.6500	2545.9282	2546.1239	-0.1958	0	K.MELDGYMVLEGATIAAYNHMNR.N 3 Oxidation (M) (Ions score 52)
176 - 199	854.6580	2560.9522	2561.1989	-0.2467	0	K.QVAMQVAAMNPIAVDEDGVSEEVK.Q 2 Oxidation (M) (Ions score 64)
290 - 303	847.2930	1692.5714	1693.7539	-1.1824	0	K.EVCLLNQEDIMDAK.K Oxidation (M) (Ions score 39)






```
LOCUS       YP_101286                330 aa            linear   BCT 26-APR-2009
DEFINITION  elongation factor Ts [Bacteroides fragilis YCH46].
ACCESSION   YP_101286
VERSION     YP_101286.1  GI:53715294
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 330)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 330)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 330)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD50752.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
```

```
source          1..330
                /organism="Bacteroides fragilis YCH46"
                /strain="YCH46"
                /db_xref="taxon:295405"
Protein         1..330
                /product="elongation factor Ts"
                /calculated_mol_wt=35884
Region          1..327
                /region_name="tsf"
                /note="elongation factor Ts; Provisional; PRK09377"
                /db_xref="CDD:137239"
Region          56..225
                /region_name="EF_TS"
                /note="Elongation factor TS; pfam00889"
                /db_xref="CDD:109927"
Region          <294..327
                /region_name="EF_TS"
                /note="Elongation factor TS; pfam00889"
                /db_xref="CDD:109927"
CDS             1..330
                /gene="tsf"
                /locus_tag="BF4010"
                /coded_by="complement(NC_006347.1:4579984..4580976)"
                /note="EF-Ts; functions during elongation stage of protein
                translation; forms a dimer; associates with EF-Tu-GDP
                complex and promotes exchange of GDP to GTP resulting in
                regeneration of the active form of EF-Tu"
                /transl_table=11
                /db_xref="GeneID:3085091"
```

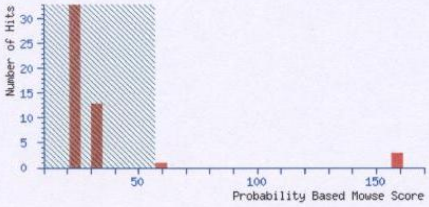
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 4705
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4705) 10_RJ18_01_1401.d\SSP (4705) 10_RJ18_01_1401.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:38:38 GMT
Protein hits    : gi|53715262  malate dehydrogenase [Bacteroides fragilis YCH46]
                  gi|224536272 hypothetical protein BACCELL_01139 [Bacteroides cellulosilyticus DSM 14838]
                  gi|160890687 hypothetical protein BACUNI_03132 [Bacteroides uniformis ATCC 8492]
                  gi|150002856 malate dehydrogenase [Bacteroides vulgatus ATCC 8482]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53715262 **Mass:** 32945 **Score:** 159 **Queries matched:** 4 **emPAI:** 0.21
malate dehydrogenase [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	121	621.2850	1240.5554	1240.7027	-0.1472	0	41	2.2	1	R.EELIGVNAGIVK.S
☑	175	853.8230	1705.6314	1705.8920	-0.2606	0	56	0.054	1	K.LNEVVASTMVGGATLTK.L + Oxidation (M)
	178	862.6630	2584.9672	2585.3121	-0.3449	0	21	2.9e+02	2	K.YSPNAIIVVISNPMDTMTYLALK.S + 2 Oxidation (M)
☑	248	906.0490	2715.1252	2715.4020	-0.2769	0	41	1.2	1	K.LLGTSAWYAPGAAGAYVVESIIHNQK.K

Proteins matching the same set of peptides:
gi|255011805 **Mass:** 32932 **Score:** 159 **Queries matched:** 4
malate dehydrogenase [Bacteroides fragilis 3_1_12]

---

2. gi|224536272 **Mass:** 33016 **Score:** 159 **Queries matched:** 4 **emPAI:** 0.21
hypothetical protein BACCELL_01139 [Bacteroides cellulosilyticus DSM 14838]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
121	621.2850	1240.5554	1240.7027	-0.1472	0	41	2.2	1	R.EELIGVNAGIVK.T
175	853.8230	1705.6314	1705.8920	-0.2606	0	56	0.054	1	K.LNEVVASTMVGGATLTK.L + Oxidation (M)
178	862.6630	2584.9672	2585.3121	-0.3449	0	21	2.9e+02	2	K.YSPNAILVVISNPMDTMTYLALK.S + 2 Oxidation (M)
248	906.0490	2715.1252	2715.4020	-0.2769	0	41	1.2	1	K.LLGTSAWYAPGAAGAYVVESIIHNQK.K

---

3. gi|160890687 **Mass:** 33051 **Score:** 159 **Queries matched:** 4 **emPAI:** 0.21
hypothetical protein BACUNI_03132 [Bacteroides uniformis ATCC 8492]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
121	621.2850	1240.5554	1240.7027	-0.1472	0	41	2.2	1	R.EELIGVNAGIVK.S
175	853.8230	1705.6314	1705.8920	-0.2606	0	56	0.054	1	K.LNEVVASTMVGGATLTK.L + Oxidation (M)
178	862.6630	2584.9672	2585.3121	-0.3449	0	20	3e+02	4	K.YSPNAILVVISNPMDTMTYLSLK.S + Oxidation (M)
248	906.0490	2715.1252	2715.4020	-0.2769	0	41	1.2	1	K.LLGTSAWYAPGAAGAYVVESIIHNQK.K

---

4. gi|150002856 **Mass:** 33174 **Score:** 62 **Queries matched:** 2
malate dehydrogenase [Bacteroides vulgatus ATCC 8482]
☐ Check to include this hit in error tolerant search

Mascot Search Results

Protein View

Match to: **gi|53715262** Score: **159**
malate dehydrogenase [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4705) 10_RJ18_01_1401.d\SSP (4705) 10_RJ18_01_1401.mgf

Nominal mass (M_r): **32945**; Calculated pI value: **5.14**
NCBI BLAST search of gi|53715262 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60683199 from Bacteroides fragilis NCTC 9343
gi|253567153 from Bacteroides sp. 3_2_5
gi|265767749 from Bacteroides sp. 2_1_16
gi|73920985 from Bacteroides fragilis NCTC 9343
gi|73920986 from Bacteroides fragilis
gi|52218127 from Bacteroides fragilis YCH46
gi|60494633 from Bacteroides fragilis NCTC 9343
gi|251943984 from Bacteroides sp. 3_2_5
gi|263252421 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **24%**

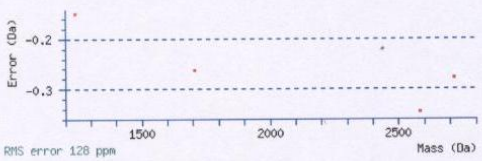
Matched peptides shown in Bold Red

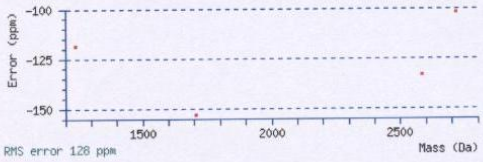
```
  1 MSKVTVVGAG NVGATCANVL AFNEVADEVV MLDVKEGVSE GKAMDMMQTA
 51 QLLGFDTTIV GCTNDYAQTA NSDVVVITSG IPRKPGMTRE ELIGVNAGIV
101 KSVAENLLKY SPNAIIVVIS NPMDTMTYLA LKSLGLPKNR VIGMGGALDS
151 SRFKYFLSQA LGCNANEVEG MVIGGHGDTT MIPLARLATY KGQPVSTLLS
201 EEKLNEVVAS TMVGGATLTK LLGTSAWYAP GAAGAYVVES IIHNQKKMVP
251 CSVMLEGEYG ESDLCIGVPV ILGKNGIEKI VELELNADEK AKFAASAAAV
301 HKTNAALKEV GAL
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
90 - 101	621.2850	1240.5554	1240.7027	-0.1472	0	R.EELIGVNAGIVK.S (Ions score 41)
110 - 132	862.6630	2584.9672	2585.3121	-0.3449	0	K.YSPNAIIVVISNPMDTMTYLALK.S 2 Oxidation (M) (Ions score 21)
204 - 220	853.8230	1705.6314	1705.8920	-0.2606	0	K.LNEVVASTMVGGATLTK.L Oxidation (M) (Ions score 56)
221 - 246	906.0490	2715.1252	2715.4020	-0.2769	0	K.LLGTSAWYAPGAAGAYVVESIIHNQK.K (Ions score 41)






```
LOCUS       YP_101254                313 aa            linear   BCT 26-APR-2009
DEFINITION  malate dehydrogenase [Bacteroides fragilis YCH46].
ACCESSION   YP_101254
VERSION     YP_101254.1  GI:53715262
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 313)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 313)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 313)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD50720.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..313
                     /organism="Bacteroides fragilis YCH46"
```

```
                /strain="YCH46"
                /db_xref="taxon:295405"
Protein         1..313
                /product="malate dehydrogenase"
                /calculated_mol_wt=32545
Region          20..286
                /region_name="LDH-like_MDH"
                /note="L-lactate dehydrogenase-like malate dehydrogenase
                proteins; cd01339"
                /db_xref="CDD:133424"
Site            order(22,37..38,41,44..45,47..48,52,54..56,151..152,154,
                156,221,227..229,232..233,236)
                /site_type="other"
                /note="dimer interface"
                /db_xref="CDD:133424"
Site            order(33..34,78..81,99,119,121,144,148,176)
                /site_type="other"
                /note="NAD(P) binding site"
                /db_xref="CDD:133424"
Site            order(54,56,162,167,169,171,189,191..193,243,245..248,274,
                283)
                /site_type="other"
                /note="tetramer (dimer of dimers) interface"
                /db_xref="CDD:133424"
Site            order(89,121,152,176,215,226)
                /site_type="other"
                /note="substrate binding site"
                /db_xref="CDD:133424"
CDS             1..313
                /locus_tag="BF3978"
                /coded_by="complement(NC_006347.1:4546190..4547131)"
                /note="similar to gp:AE016942_184 [Bacteroides
                thetaiotaomicron VPI-5482], percent identity 92 in 313 aa,
                BLASTP E(): e-163"
                /transl_table=11
                /db_xref="GeneID:3085059"
```

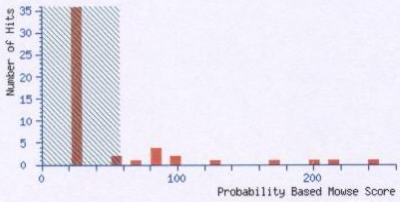
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User          : Lakshmy Manickan
Email         : lakshmy.manickan@unn.ac.uk
Search title  : SSP 4804
MS data file  : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4804) 100_RE5_01_1707.d\SSP (4804) 100_RE5_01_1707.mgf
Database      : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy      : Bacteria (Eubacteria) (5690016 sequences)
Timestamp     : 6 Jan 2010 at 13:39:46 GMT
Protein hits  : gi|223043068 pyridoxine biosynthesis protein [Staphylococcus capitis SK14]
                gi|228475242 pyridoxine biosynthesis protein [Staphylococcus hominis SK119]
                gi|15923509  pyridoxal biosynthesis lyase PdxS [Staphylococcus aureus subsp. aureus Mu50]
                gi|53715262  malate dehydrogenase [Bacteroides fragilis YCH46]
                gi|167765305 hypothetical protein BACSTE_03693 [Bacteroides stercoris ATCC 43183]
                gi|224536272 hypothetical protein BACCELL_01139 [Bacteroides cellulosilyticus DSM 14838]
                gi|160890687 hypothetical protein BACUNI_03132 [Bacteroides uniformis ATCC 8492]
                gi|229542250 pyridoxine biosynthesis protein [Bacillus coagulans 36D1]
                gi|23100142  pyridoxine biosynthesis protein [Oceanobacillus iheyensis HTE831]
                gi|160947252 hypothetical protein PEPMIC_01185 [Parvimonas micra ATCC 33270]
                gi|239617197 pyridoxine biosynthesis protein [Kosmotoga olearia TBF 19.5.1]
                gi|83592546  malate dehydrogenase [Rhodospirillum rubrum ATCC 11170]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|223043068 **Mass:** 32030 **Score:** 244 **Queries matched:** 5 **emPAI:** 0.48
pyridoxine biosynthesis protein [Staphylococcus capitis SK14]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	155	693.7620	1385.5094	1385.7150	-0.2056	0	55	0.083	1	K.GLDINQISLEER.M
☑	169	773.3010	1544.5874	1544.7868	-0.1994	0	83	0.00013	1	K.IAEEAGAVAVMALER.V + Oxidation (M)
☑	179	830.2580	1658.5014	1658.7532	-0.2517	0	64	0.0068	1	R.LTVMNDDEIMTFAK.D + 2 Oxidation (M)
	202	**650.5870**	**1948.7392**	**1948.9894**	**-0.2503**	**0**	**(21)**	**1.4e+02**	**2**	**K.AIVQATTHYQDYELIGK.L**
☑	203	975.4180	1948.8214	1948.9894	-0.1680	0	42	1.4	1	K.AIVQATTHYQDYELIGK.L

Proteins matching the same set of peptides:
gi|239636916 **Mass:** 31970 **Score:** 244 **Queries matched:** 5
pyridoxine biosynthesis protein [Staphylococcus warneri L37603]
gi|242372735 **Mass:** 32072 **Score:** 244 **Queries matched:** 5
pyridoxine (vitamin B6) biosynthesis protein [Staphylococcus epidermidis M23864:W1]
gi|27469180 **Mass:** 32048 **Score:** 242 **Queries matched:** 5
pyridoxine biosynthesis protein [Staphylococcus epidermidis ATCC 12228]
gi|251809935 **Mass:** 32038 **Score:** 242 **Queries matched:** 5
pyridoxine (vitamin B6) biosynthesis protein [Staphylococcus epidermidis BCM-HMP0060]

---

2. gi|228475242 **Mass:** 32014 **Score:** 215 **Queries matched:** 5 **emPAI:** 0.35
pyridoxine biosynthesis protein [Staphylococcus hominis SK119]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
155	693.7620	1385.5094	1385.7514	-0.2420	1	55	0.083	1	K.GLDINKISLEER.M
169	773.3010	1544.5874	1544.7868	-0.1994	0	83	0.00013	1	R.IAEEAGAVAVMALER.V + Oxidation (M)
179	830.2580	1658.5014	1658.7531	-0.2517	0	35	5.4	2	R.LTVMNDDEIMTYAK.E + Oxidation (M)
202	650.5870	1948.7392	1948.9894	-0.2503	0	(21)	1.4e+02	2	K.AIVQATTHYQDYELIGK.L
203	975.4180	1948.8214	1948.9894	-0.1680	0	42	1.4	1	K.AIVQATTHYQDYELIGK.L

---

3. gi|15923509 **Mass:** 32088 **Score:** 202 **Queries matched:** 3 **emPAI:** 0.34
pyridoxal biosynthesis lyase PdxS [Staphylococcus aureus subsp. aureus Mu50]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
155	693.7620	1385.5094	1385.7150	-0.2056	0	55	0.083	1	K.GLDINQLSLEER.M
169	773.3010	1544.5874	1544.7868	-0.1994	0	83	0.00013	1	R.IAEEAGAVAVMALER.V + Oxidation (M)
179	830.2580	1658.5014	1658.7532	-0.2517	0	64	0.0068	1	R.LTVMNDDEIMTFAK.D + 2 Oxidation (M)

Proteins matching the same set of peptides:
gi|82750226 **Mass:** 31832 **Score:** 202 **Queries matched:** 3
pyridoxine biosynthesis protein [Staphylococcus aureus RF122]

---

4. gi|53715262 **Mass:** 32945 **Score:** 171 **Queries matched:** 5 **emPAI:** 0.21
malate dehydrogenase [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 136	621.2730	1240.5314	1240.7027	-0.1712	0	33	15	1	R.EELIGVNAGIVK.S
☑ 141	636.7330	1271.4514	1271.6609	-0.2094	0	78	0.0004	1	K.IVELELNADEK.A
☑ 247	862.6770	2585.0092	2585.3121	-0.3029	0	21	1.2e+02	1	K.YSPNAIIVVISNPMDTMTYLALK.S + 2 Oxidation (M)
☑ 250	992.4190	2974.2352	2974.5533	-0.3181	1	40	1.6	1	K.GQPVSTLLSEEKLNEVVASTMVGGATLTK.L + Oxidation (M)
251	**992.4990**	**2974.4752**	**2974.5533**	**-0.0781**	**1**	**(17)**	**3.6e+02**	**2**	**K.GQPVSTLLSEEKLNEVVASTMVGGATLTK.L + Oxidation (M)**

---

5. gi|167765305 **Mass:** 33007 **Score:** 130 **Queries matched:** 3 **emPAI:** 0.10
hypothetical protein BACSTE_03693 [Bacteroides stercoris ATCC 43183]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
136	621.2730	1240.5314	1240.7027	-0.1712	0	33	15	1	R.EELIGVNAGIVK.S
141	636.7330	1271.4514	1271.6609	-0.2094	0	78	0.0004	1	K.IVELELNADEK.A
247	862.6770	2585.0092	2585.3121	-0.3029	0	20	1.6e+02	3	K.YSPNAILVVISNPMDTMTYLSLK.S + Oxidation (M)

Proteins matching the same set of peptides:
gi|218128506 **Mass:** 32901 **Score:** 130 **Queries matched:** 3
hypothetical protein BACEGG_00076 [Bacteroides eggerthii DSM 20697]

---

6. gi|224536272 **Mass:** 33016 **Score:** 93 **Queries matched:** 4 **emPAI:** 0.10
hypothetical protein BACCELL_01139 [Bacteroides cellulosilyticus DSM 14838]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
136	621.2730	1240.5314	1240.7027	-0.1712	0	33	15	1	R.EELIGVNAGIVK.T
247	862.6770	2585.0092	2585.3121	-0.3029	0	21	1.2e+02	1	K.YSPNAILVVISNPMDTMTYLALK.S + 2 Oxidation (M)
250	992.4190	2974.2352	2974.5533	-0.3181	1	40	1.6	1	K.GQPVSTLLSEEKLNEVVASTMVGGATLTK.L + Oxidation (M)
251	992.4990	2974.4752	2974.5533	-0.0781	1	(17)	3.6e+02	2	K.GQPVSTLLSEEKLNEVVASTMVGGATLTK.L + Oxidation (M)

---

7. gi|160890687 **Mass:** 33051 **Score:** 92 **Queries matched:** 4 **emPAI:** 0.10
hypothetical protein BACUNI_03132 [Bacteroides uniformis ATCC 8492]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
136	621.2730	1240.5314	1240.7027	-0.1712	0	33	15	1	R.EELIGVNAGIVK.S
247	862.6770	2585.0092	2585.3121	-0.3029	0	20	1.6e+02	3	K.YSPNAILVVISNPMDTMTYLSLK.S + Oxidation (M)
250	992.4190	2974.2352	2974.5533	-0.3181	1	40	1.6	1	K.GQPVSTLLSEEKLNEVVASTMVGGATLTK.L + Oxidation (M)
251	992.4990	2974.4752	2974.5533	-0.0781	1	(17)	3.6e+02	2	K.GQPVSTLLSEEKLNEVVASTMVGGATLTK.L + Oxidation (M)

---

8. gi|229542250 **Score:** 83 **Queries matched:** 1
pyridoxine biosynthesis protein [Bacillus coagulans 36D1]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
169	773.3010	1544.5874	1544.7868	-0.1994	0	83	0.00013	1	K.LAEEAGAVAVMALER.V + Oxidation (M)

---

9. gi|23100142 **Score:** 82 **Queries matched:** 1
pyridoxine biosynthesis protein [Oceanobacillus iheyensis HTE831]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
169	773.3010	1544.5874	1543.8028	0.7847	0	82	0.00013	3	K.IAEQAGAVAVMALER.V + Oxidation (M)

---

10. gi|160947252 **Score:** 82 **Queries matched:** 1
hypothetical protein PEPMIC_01185 [Parvimonas micra ATCC 33270]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
169	773.3010	1544.5874	1543.8392	0.7483	1	82	0.00013	3	K.IAEKAGAVAVMALER.I + Oxidation (M)

---

11. gi|239617197 **Score:** 80 **Queries matched:** 1

Mascot Search Results

Protein View

Match to: **gi|53715262** Score: **171**
malate dehydrogenase [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (4804) 100_RE5_01_1707.d\SSP (4804) 100_RE5_01_1707.mgf

Nominal mass (M_r): **32945**; Calculated pI value: **5.14**
NCBI BLAST search of gi|53715262 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60683199 from Bacteroides fragilis NCTC 9343
gi|253567153 from Bacteroides sp. 3_2_5
gi|265767749 from Bacteroides sp. 2_1_16
gi|73920985 from Bacteroides fragilis NCTC 9343
gi|73920986 from Bacteroides fragilis
gi|52218127 from Bacteroides fragilis YCH46
gi|60494633 from Bacteroides fragilis NCTC 9343
gi|251943984 from Bacteroides sp. 3_2_5
gi|263252421 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **23%**

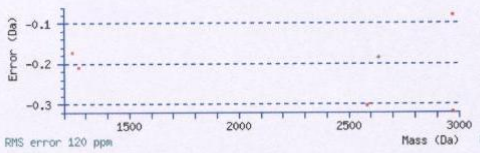
Matched peptides shown in Bold Red

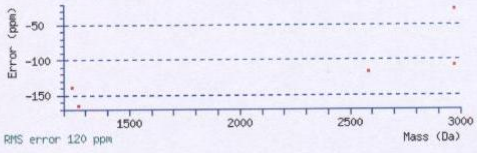
```
  1 MSKVTVVGAG NVGATCANVL AFNEVADEVV MLDVKEGVSE GKAMDMMQTA
 51 QLLGFDTTIV GCTNDYAQTA NSDVVVITSG IPRKPGMTRE ELIGVNAGIV
101 KSVAENLLKY SPNAIIVVIS NPMDTMTYLA LKSLGLPKNR VIGMGGALDS
151 SRFKYFLSQA LGCNANEVEG MVIGGHGDTT MIPLARLATY KGQPVSTLLS
201 EEKLNEVVAS TMVGGATLTK LLGTSAWYAP GAAGAYVVES IIHNQKKMVP
251 CSVMLEGEYG ESDLCIGVPV ILGKNGIEKI VELELNADEK AKFAASAAAV
301 HKTNAALKEV GAL
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
90 - 101	621.2730	1240.5314	1240.7027	-0.1712	0	R.EELIGVNAGIVK.S (Ions score 33)
110 - 132	862.6770	2585.0092	2585.3121	-0.3029	0	K.YSPNAIIVVISNPMDTMTYLALK.S 2 Oxidation (M) (Ions score 21)
192 - 220	992.4190	2974.2352	2974.5533	-0.3181	1	K.GQPVSTLLSEEKLNEVVASTMVGGATLTK.L Oxidation (M) (Ions score 40)
192 - 220	992.4990	2974.4752	2974.5533	-0.0781	1	K.GQPVSTLLSEEKLNEVVASTMVGGATLTK.L Oxidation (M) (Ions score 17)
280 - 290	636.7330	1271.4514	1271.6609	-0.2094	0	K.IVELELNADEK.A (Ions score 78)






```
LOCUS       YP_101254                313 aa            linear   BCT 26-APR-2009
DEFINITION  malate dehydrogenase [Bacteroides fragilis YCH46].
ACCESSION   YP_101254
VERSION     YP_101254.1  GI:53715262
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 313)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 313)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 313)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD50720.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..313
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
```

```
Protein         1..313
                /product="malate dehydrogenase"
                /calculated_mol_wt=32545
Region          20..286
                /region_name="LDH-like_MDH"
                /note="L-lactate dehydrogenase-like malate dehydrogenase
                proteins; cd01339"
                /db_xref="CDD:133424"
Site            order(22,37..38,41,44..45,47..48,52,54..56,151..152,154,
                156,221,227..229,232..233,236)
                /site_type="other"
                /note="dimer interface"
                /db_xref="CDD:133424"
Site            order(33..34,78..81,99,119,121,144,148,176)
                /site_type="other"
                /note="NAD(P) binding site"
                /db_xref="CDD:133424"
Site            order(54,56,162,167,169,171,189,191..193,243,245..248,274,
                283)
                /site_type="other"
                /note="tetramer (dimer of dimers) interface"
                /db_xref="CDD:133424"
Site            order(89,121,152,176,215,226)
                /site_type="other"
                /note="substrate binding site"
                /db_xref="CDD:133424"
CDS             1..313
                /locus_tag="BF3978"
                /coded_by="complement(NC_006347.1:4546190..4547131)"
                /note="similar to gp:AE016942_184 [Bacteroides
                thetaiotaomicron VPI-5482], percent identity 92 in 313 aa,
                BLASTP E(): e-163"
                /transl_table=11
                /db_xref="GeneID:3085059"
```

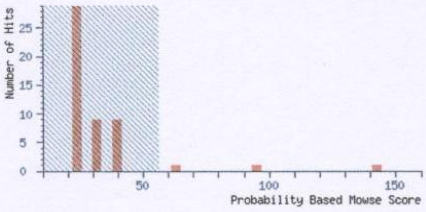
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : LAKSHMY MANICKAN
Email           : lakshmy.manickan@unn.ac.uk
Search title    : 5106 bf
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (5106) 10_RI5_01_1306.d\SSP (5106) 10_RI5_01_1306.mgf
Database        : NCBInr 20090528 (8915381 sequences; 3049521622 residues)
Taxonomy        : Bacteria (Eubacteria) (4801653 sequences)
Timestamp       : 6 Jun 2009 at 11:30:52 GMT
Protein hits    : gi|126496    RecName: Full=Lysostaphin; AltName: Full=Glycyl-glycine endopeptidase; Flags: Precursor
                  gi|29348144 50S ribosomal protein L7/L12 [Bacteroides thetaiotaomicron VPI-5482]
                  gi|53713908 hypothetical protein BF2616 [Bacteroides fragilis YCH46]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 56 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ Error tolerant

1. gi|126496 **Mass:** 51656 **Score:** 142 **Queries matched:** 3 **emPAI:** 0.13
RecName: Full=Lysostaphin; AltName: Full=Glycyl-glycine endopeptidase; Flags: Precursor
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 104	819.8400	1637.6654	1637.7897	-0.1242	0	57	0.044	1	K.SESASFTPNTDIITR.T
☑ 166	1023.3990	2044.7834	2044.9234	-0.1400	0	86	4.2e-05	1	R.MVNSFSNSTAQDPMPFLK.S + 2 Oxidation (M)
☑ 167	682.6050	2044.7932	2044.9234	-0.1302	0	(30)	16	1	R.MVNSFSNSTAQDPMPFLK.S + 2 Oxidation (M)

Proteins matching the same set of peptides:
gi|153047 **Mass:** 42213 **Score:** 142 **Queries matched:** 3
lysostaphin (ttg start codon)
gi|3287967 **Mass:** 53058 **Score:** 142 **Queries matched:** 3
RecName: Full=Lysostaphin; AltName: Full=Glycyl-glycine endopeptidase; Flags: Precursor

---

2. gi|29348144 **Mass:** 12730 **Score:** 93 **Queries matched:** 1 **emPAI:** 0.27
50S ribosomal protein L7/L12 [Bacteroides thetaiotaomicron VPI-5482]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 83	666.8460	1331.6774	1331.7449	-0.0674	0	93	1.2e-05	1	K.AFAEQLVNLTVK.E

Proteins matching the same set of peptides:
gi|53715478 **Mass:** 12688 **Score:** 93 **Queries matched:** 1
50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46]
gi|110639543 **Mass:** 12652 **Score:** 93 **Queries matched:** 1
50S ribosomal protein L7/L12 [Cytophaga hutchinsonii ATCC 33406]
gi|149278891 **Mass:** 12813 **Score:** 93 **Queries matched:** 1
50S ribosomal protein L7/L12 [Pedobacter sp. BAL39]
gi|150003396 **Mass:** 12704 **Score:** 93 **Queries matched:** 1
50S ribosomal protein L7/L12 [Bacteroides vulgatus ATCC 8482]
gi|150008875 **Mass:** 12727 **Score:** 93 **Queries matched:** 1
50S ribosomal protein L7/L12 [Parabacteroides distasonis ATCC 8503]
gi|153805949 **Mass:** 12686 **Score:** 93 **Queries matched:** 1
hypothetical protein BACCAC_00194 [Bacteroides caccae ATCC 43185]
gi|154491722 **Mass:** 12679 **Score:** 93 **Queries matched:** 1
hypothetical protein PARMER_01333 [Parabacteroides merdae ATCC 43184]
gi|160888407 **Mass:** 12775 **Score:** 93 **Queries matched:** 1

hypothetical protein BACUNI_00820 [Bacteroides uniformis ATCC 8492]
gi|189460676 **Mass:** 12771 **Score:** 93 **Queries matched:** 1
hypothetical protein BACCOP_01323 [Bacteroides coprocola DSM 17136]
gi|189465410 **Mass:** 12759 **Score:** 93 **Queries matched:** 1
hypothetical protein BACINT_01763 [Bacteroides intestinalis DSM 17393]
gi|198277435 **Mass:** 12757 **Score:** 93 **Queries matched:** 1
hypothetical protein BACPLE_03654 [Bacteroides plebeius DSM 17135]
gi|224023599 **Mass:** 12771 **Score:** 93 **Queries matched:** 1
hypothetical protein BACCOPRO_00306 [Bacteroides coprophilus DSM 18228]
gi|224537808 **Mass:** 12757 **Score:** 93 **Queries matched:** 1
hypothetical protein BACCELL_02695 [Bacteroides cellulosilyticus DSM 14838]
gi|227413422 **Mass:** 13015 **Score:** 93 **Queries matched:** 1
LSU ribosomal protein L12P [Dyadobacter fermentans DSM 18053]

---

3. gi|53713908 **Mass:** 12967 **Score:** 60 **Queries matched:** 3
hypothetical protein BF2616 [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	120	590.5810	1768.7212	1768.9723	-0.2511	0	33	8.8	1	K.LPKPNLIAAQEYVEGK.I
☑	121	885.4150	1768.8154	1768.9723	-0.1568	0	(17)	4.9e+02	1	K.LPKPNLIAAQEYVEGK.I
☑	219	763.3310	2286.9712	2287.0756	-0.1045	1	28	25	1	K.KLAEFASYNDNLEEVFENR.E

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	88	681.0850	2040.2332	2040.0528	0.1804	1	39	4	1	EVFVSGLREYLGDTAGSLK
☑	108	559.9450	1676.8132	1676.8958	-0.0827	1	37	4.7	1	DGVVAVVRSHSPDLAR
☑	179	1041.9630	2081.9114	2080.9955	0.9159	1	36	5	1	TMREEIDAMEVMGVSVIR + Oxidation (M)
☑	109	843.7830	2528.3272	2528.2693	0.0579	1	30	33	1	IWSAGVAASPLGKMVAEQAGVEADR + Oxidation (M)
☑	204	737.6530	2209.9372	2211.0598	-1.1227	0	29	23	1	EHPRPPVMMSTASQMLVER + Oxidation (M)
☑	209	742.3440	2224.0102	2225.2063	-1.1961	2	28	32	1	MLLSLPVQQGEGGVNLSRRR + Oxidation (M)
☑	205	1105.9790	2209.9434	2210.1306	-0.1872	1	28	30	1	CQIVYNVLKTHIQFGAYR
☑	243	929.9330	2786.7772	2787.4627	-0.6855	1	26	30	1	HSIPLTAALALGAGAARAAQAPTSTDEAR
☑	236	878.4340	2632.2802	2633.3517	-1.0715	1	26	48	1	MDLNLQDKVVIVTGGASGIGGAISMR + 2 Oxidation (M)
☑	37	819.1320	818.1247	817.4480	0.6767	1	26	55	1	MGIDKVR
☑	202	731.8270	2192.4592	2192.1114	0.3478	0	26	39	1	QLDPSGAVGWPVVAALADGLGAE
☑	70	572.2670	1713.7792	1713.8210	-0.0418	0	25	1.6e+02	1	DIIEHPEEQSGGYIK
☑	231	834.3320	2499.9742	2499.3261	0.6481	1	25	43	1	LTAEQVLDKTIVLFNYVYENK
☑	247	968.7290	2903.1652	2903.5604	-0.3952	2	24	42	1	RARLFQDEASVEDLATILTIILSSDK
☑	184	700.9170	2099.7292	2100.1263	-0.3971	1	24	52	1	GGFRHLIVTSGGGVAGILSMR + Oxidation (M)
☑	126	608.5840	1822.7302	1821.8508	0.8793	1	24	68	1	FAEGWMARVGYGTAYK + Oxidation (M)
☑	163	675.9900	2024.9482	2026.0769	-1.1287	0	24	86	1	GAVINMTGPSVDVNLGALTVV
☑	172	688.2020	2061.5842	2062.0272	-0.4431	1	24	61	1	WPGSDYDGLWVVTIRNGK
☑	178	694.9630	2081.8672	2081.0800	0.7872	2	24	78	1	GHDLSGRRITVSTCGIVPR
☑	235	866.7260	2597.1562	2597.2386	-0.0825	2	23	79	1	MIAKDGEGATKLIECTVMNVSGEK + Oxidation (M)
☑	73	592.7900	1183.5654	1184.6765	-1.1110	1	23	1.2e+02	1	LKPTGNKVTLD
☑	171	688.1910	2061.5512	2062.0259	-0.4747	0	23	70	1	IDVLNNSLIGSTFSYEYK
☑	240	914.1200	2739.3382	2739.3717	-0.0335	0	23	88	1	YFIVPEYHILTLTPDSISMEDLK + Oxidation (M)
☑	146	647.6210	1939.8412	1940.0699	-0.2287	1	23	1e+02	1	LRLLANVLPAMLDQMAR + Oxidation (M)
☑	139	945.0180	2832.0322	2831.4424	0.5898	1	23	1.7e+02	1	MEIAALVIVMGMQFAVYAPIYKADR + 2 Oxidation (M)
☑	246	937.0270	2808.0592	2807.4449	0.6143	1	23	62	1	TSAITGWKIVGVSPEAEMTSGVIMLSK + Oxidation (M)
☑	203	1103.5350	3307.5832	3307.7346	-0.1514	2	22	1.9e+02	1	VLNAKYHEKEAEIIAQAGAPGALTIATNMAGR
☑	220	767.3070	2298.8992	2299.1193	-0.2201	1	22	83	1	KVSNEWHGNLTVSGNVTESNK
☑	248	978.0340	2931.0802	2930.4670	0.6132	2	22	67	1	LKADCDSPLKVYAHISPYMGGPEIIR
☑	218	1130.4240	3388.2502	3387.5509	0.6993	2	22	1.4e+02	1	EQTEVPMVMWMSPGFAASFGVNKDCLQRR + Oxidation (M)
☑	227	813.3450	2437.0132	2438.1060	-1.0928	0	22	1e+02	1	GFGFVEMPNDEEGNAAIAALNEK + Oxidation (M)
☑	148	976.3900	2926.1482	2926.4519	-0.3038	1	22	1.9e+02	1	SRDIDEGLALALAAAADLTGDDLEAGLDR
☑	113	864.0980	2589.2722	2588.3996	0.8726	1	21	2.5e+02	1	LGMGKPLGLGSVAINYELQLTDRK + Oxidation (M)
☑	140	632.7910	1895.3512	1894.0447	0.5064	2	21	1.1e+02	1	SADRANDMNGMMRDIAK
☑	180	695.6620	2083.9642	2083.9844	-0.0203	1	21	1.5e+02	1	DNIGKMLSFTTENGQGQTK + Oxidation (M)
☑	101	804.1330	2409.3772	2409.3168	0.0603	2	21	2.6e+02	1	LVELVEAALPGARDGWLYPGRK
☑	154	663.5160	1987.5262	1988.0626	-0.5364	2	21	1.2e+02	1	VRVDYANGVLQVMGPKAR + Oxidation (M)
☑	102	809.7900	2426.3482	2425.3428	1.0054	2	21	2.8e+02	1	ANIIAKSGKAVEVAGDIDTVVLDK
☑	250	989.7350	2966.1832	2966.5059	-0.3227	2	21	93	1	NSDYVIFVQNKIYSVSLNKSLMSSSK + Oxidation (M)
☑	133	618.7790	1853.3152	1853.8465	-0.5313	0	21	1.2e+02	1	LESVLDEAEAMANSFGR + Oxidation (M)
☑	196	1081.5960	3241.7662	3240.8632	0.9030	2	21	2.8e+02	1	MTLDKLGSAFLLGIALASLGVPALARGISTLR + Oxidation (M)
☑	74	623.6370	1867.8892	1866.9112	0.9780	0	21	3.3e+02	1	YLNDPSYAGSIIGPTSGR
☑	115	874.7990	2621.3752	2621.3418	0.0334	1	21	2.9e+02	1	GMEVAAAGGHNLLMTGPPGAGKTLLAR + 2 Oxidation (M)
☑	221	771.7050	2312.0932	2312.0590	0.0341	1	21	1.7e+02	1	TADLGSCSRQDIVNAYDNSLL
☑	190	704.7750	2111.3032	2111.0371	0.2661	1	21	1.4e+02	1	FSHFMGAVVLEARGGFSAGR + Oxidation (M)
☑	149	656.6840	1967.0302	1966.9935	0.0367	1	21	1.9e+02	1	MPDGKHWVIGTVIDAQGK + Oxidation (M)
☑	194	1075.9480	3224.8222	3224.6032	0.2190	0	20	2.7e+02	1	MNGLAYSGAVLATAVGGFDPAPMIIMAAALGAR + 3 Oxidation (M)
☑	75	624.5210	1870.5412	1870.9398	-0.3986	2	20	3.3e+02	1	DFGATHEELAGRLGRSR
☑	206	1106.2650	3315.7732	3315.5621	0.2111	1	20	2.8e+02	1	GQGIANIVNGFFGGMAGCAMIGQSMINVKSGGR + Oxidation (M)
☑	111	854.3570	1706.6994	1706.8125	-0.1130	1	20	2.1e+02	1	NARYASSFDVGEHVR
☑	71	581.2250	1740.6532	1741.7968	-1.1436	0	20	4.4e+02	1	ILDDYIEMEEELSK + Oxidation (M)
☑	197	1082.6650	3244.9732	3245.5746	-0.6014	1	20	2.8e+02	1	VAGVGDAMVMGVDYSMRIWLKPDVMAQYK + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|53715478** Score: **93**
50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (5106) 10_RI5_01_1306.d\SSP (5106) 10_RI5_01_1306.mgf

Nominal mass (M_r): **12688**; Calculated pI value: **4.71**
NCBI BLAST search of gi|53715478 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60683451 from Bacteroides fragilis NCTC 9343
gi|253566657 from Bacteroides sp. 3_2_5
gi|255011634 from Bacteroides fragilis 3_1_12
gi|265767535 from Bacteroides sp. 2_1_16
gi|81313501 from Bacteroides fragilis NCTC 9343
gi|81608231 from Bacteroides fragilis
gi|52218343 from Bacteroides fragilis YCH46
gi|60494885 from Bacteroides fragilis NCTC 9343
gi|251944829 from Bacteroides sp. 3_2_5
gi|263252840 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **9%**

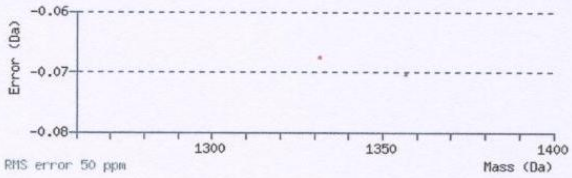
Matched peptides shown in Bold Red

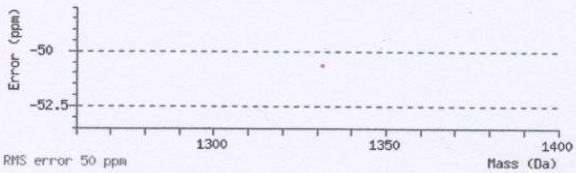
```
  1 MADLKAFAEQ LVNLTVKEVN ELATILKEEY GIEPAAAAVA VAAGPAAGAA
 51 AAEEKSSFDV VLKSAGAAKL QVVKAVKEAC GLGLKEAKDM VDGAPSVVKE
101 GLAKDEAESL KKTLEEAGAE VELK
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
6 - 17	666.8460	1331.6774	1331.7449	-0.0674	0	K.AFAEQLVNLTVK.E (Ions score 93)






```
LOCUS       YP_101470                124 aa            linear   BCT 26-APR-2009
DEFINITION  50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46].
ACCESSION   YP_101470
VERSION     YP_101470.1  GI:53715478
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 124)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 124)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 124)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
```

```
            reference sequence was derived from BAD50936.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..124
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..124
                     /product="50S ribosomal protein L7/L12"
                     /calculated_mol_wt=12506
     Region          2..123
                     /region_name="Ribosomal_L7_L12"
                     /note="Ribosomal protein L7/L12. Ribosomal protein L7/L12
                     refers to the large ribosomal subunit proteins L7 and L12,
                     which are identical except that L7 is acetylated at the N
                     terminus. It is a component of the L7/L12 stalk, which is
                     located at the surface of...; cd00387"
                     /db_xref="CDD:100102"
     Site            order(2,15..17,20..21,24,27,42,45..46,48..50,56..57,78,
                     80..81,83,85..86,100..101,103,106)
                     /site_type="other"
                     /note="core dimer interface"
                     /db_xref="CDD:100102"
     Site            order(5,8,12,25,30)
                     /site_type="other"
                     /note="peripheral dimer interface"
                     /db_xref="CDD:100102"
     Site            order(19,22..23,27,30)
                     /site_type="other"
                     /note="L10 interface"
                     /db_xref="CDD:100102"
     Site            order(69..70,73..74,77,84..85,88)
                     /site_type="other"
                     /note="L11 interface"
                     /db_xref="CDD:100102"
     Site            order(70,74,77,84..85,88)
                     /site_type="other"
                     /note="putative EF-Tu interaction site"
                     /db_xref="CDD:100102"
     Site            order(70,73..74,77)
                     /site_type="other"
                     /note="putative EF-G interaction site"
                     /db_xref="CDD:100102"
     CDS             1..124
                     /gene="rplL"
                     /locus_tag="BF4193"
                     /coded_by="complement(NC_006347.1:4790804..4791178)"
                     /note="present in two forms; L12 is normal, while L7 is
                     aminoacylated at the N-terminal serine; the only multicopy
                     ribosomal protein, 4:1 ratio of L7/L12 per ribosome; two
                     L12 dimers bind L10; critically important for translation
                     efficiency and fidelity; stimulates GTPase activity of
                     translation factors"
                     /transl_table=11
                     /db_xref="GeneID:3085275"
```

Mascot: http://www.matrixscience.com/

Mascot Search Results

Protein View

Match to: **gi|53713908** Score: **60**
hypothetical protein BF2616 [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (5106) 10_RI5_01_1306.d\SSP (5106) 10_RI5_01_1306.mgf

Nominal mass (M_r): **12967**; Calculated pI value: **5.28**
NCBI BLAST search of gi|53713908 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60682117 from Bacteroides fragilis NCTC 9343
gi|253567099 from Bacteroides sp. 3_2_5
gi|265764254 from Bacteroides sp. 2_1_16
gi|52216773 from Bacteroides fragilis YCH46
gi|60493551 from Bacteroides fragilis NCTC 9343
gi|251944223 from Bacteroides sp. 3_2_5
gi|263256862 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **31%**

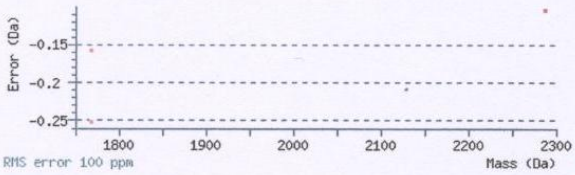
Matched peptides shown in Bold Red

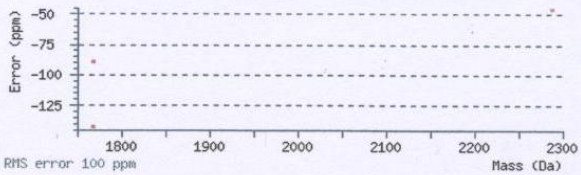
```
  1 MDYKKTNAPT NTITRDMMDL CADTGNVYET VAIIGKRANQ ISVEIKNDLS
 51 KKLAEFASYN DNLEEVFENR EQIEISRYYE KLPKPNLIAA QEYVEGKIYY
101 RNPAKEKEKL Q
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence	
52 - 70	763.3310	2286.9712	2287.0756	-0.1045	1	K.KLAEFASYNDNLEEVFENR.E	(Ions score 28)
82 - 97	590.5810	1768.7212	1768.9723	-0.2511	0	K.LPKPNLIAAQEYVEGK.I	(Ions score 33)
82 - 97	885.4150	1768.8154	1768.9723	-0.1568	0	K.LPKPNLIAAQEYVEGK.I	(Ions score 17)






```
LOCUS       YP_099900                111 aa            linear   BCT 26-APR-2009
DEFINITION  hypothetical protein BF2616 [Bacteroides fragilis YCH46].
ACCESSION   YP_099900
VERSION     YP_099900.1  GI:53713908
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 111)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 111)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 111)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD49366.
```

```
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..111
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..111
                     /product="hypothetical protein"
                     /calculated_mol_wt=12785
     CDS             1..111
                     /locus_tag="BF2616"
                     /coded_by="complement(NC_006347.1:3008816..3009151)"
                     /note="similar to gp:AE016928_87 [Bacteroides
                     thetaiotaomicron VPI-5482], percent identity 93 in 111 aa,
                     BLASTP E(): 3e-55"
                     /transl_table=11
                     /db_xref="GeneID:3083721"
```

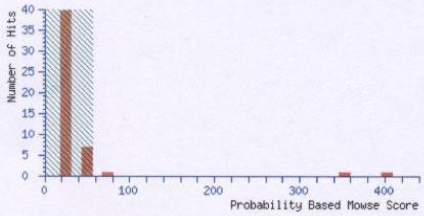
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 5205
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (5205) 10_RJ7_01_1371.d\SSP (5205) 10_RJ7_01_1371.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:41:02 GMT
Protein hits    : gi|53712495  hypothetical protein BF1203 [Bacteroides fragilis YCH46]
                  gi|255007972 hypothetical protein Bfra3_02458 [Bacteroides fragilis 3_1_12]
                  gi|153807766 hypothetical protein BACCAC_02049 [Bacteroides caccae ATCC 43185]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53712495 **Mass:** 15517 **Score:** 403 **Queries matched:** 7 **emPAI:** 1.69
hypothetical protein BF1203 [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	139	667.2930	1332.5714	1332.6959	-0.1244	0	49	0.35	1	K.NMLIVETIDAAK.K + Oxidation (M)
☑	146	688.8010	1375.5874	1375.8075	-0.2200	0	40	4.6	1	K.TILSISGKPGLYK.L
☑	159	733.7710	1465.5274	1465.7089	-0.1814	0	56	0.054	1	R.EYLAEVLPDFDR.D
	186	**579.8860**	**1736.6362**	**1736.8369**	**-0.2008**	**1**	**37**	**3.8**	**2**	**R.EYLAEVLPDFDRDR.V**
☑	224	712.9460	2135.8162	2136.0772	-0.2611	0	(60)	0.017	1	K.IISLADIAMYTNDSEVPLR.D + Oxidation (M)
☑	225	1068.9160	2135.8174	2136.0772	-0.2598	0	102	1.1e-06	1	K.IISLADIAMYTNDSEVPLR.D + Oxidation (M)
☑	248	874.0220	2619.0442	2619.3578	-0.3136	1	119	2e-08	1	K.IISLADIAMYTNDSEVPLRDVLR.S + Oxidation (M)

---

2. gi|255007972 **Mass:** 15498 **Score:** 352 **Queries matched:** 6 **emPAI:** 1.69
hypothetical protein Bfra3_02458 [Bacteroides fragilis 3_1_12]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	146	688.8010	1375.5874	1375.8075	-0.2200	0	40	4.6	1	K.TILSISGKPGLYK.L
	159	733.7710	1465.5274	1464.7249	0.8026	0	56	0.055	2	R.QYLAEVLPDFDR.D
☑	186	579.8860	1736.6362	1735.8529	0.7832	1	37	3.7	1	R.QYLAEVLPDFDRDR.V
	224	712.9460	2135.8162	2136.0772	-0.2611	0	(60)	0.017	1	K.IISLADIAMYTNDSEVPLR.D + Oxidation (M)
	225	1068.9160	2135.8174	2136.0772	-0.2598	0	102	1.1e-06	1	K.IISLADIAMYTNDSEVPLR.D + Oxidation (M)
	248	874.0220	2619.0442	2619.3578	-0.3136	1	119	2e-08	1	K.IISLADIAMYTNDSEVPLRDVLR.S + Oxidation (M)

---

3. gi|153807766 **Mass:** 16006 **Score:** 77 **Queries matched:** 2 **emPAI:** 0.21
hypothetical protein BACCAC_02049 [Bacteroides caccae ATCC 43185]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
146	688.8010	1375.5874	1375.8075	-0.2200	0	40	4.6	1	K.TILSISGKPGLYK.L
186	579.8860	1736.6362	1735.8529	0.7833	1	37	4.2	3	R.EYLAEVLPNFDRDR.V

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	150	695.1280	1388.2414	1388.6969	-0.4555	1	35	6.5	1	MKANLTLDGDPAK + Oxidation (M)
	227	713.9140	2138.7202	2139.0194	-0.2992	0	34	6.4	1	MYQELSEDPPVIFLNNSK + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|53712495** Score: **403**
hypothetical protein BF1203 [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (5205) 10_RJ7_01_1371.d\SSP (5205) 10_RJ7_01_1371.mgf

Nominal mass (M_r): **15517**; Calculated pI value: **5.15**
NCBI BLAST search of gi|53712495 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60680695 from Bacteroides fragilis NCTC 9343
gi|253563474 from Bacteroides sp. 3_2_5
gi|265762658 from Bacteroides sp. 2_1_16
gi|52215360 from Bacteroides fragilis YCH46
gi|60492129 from Bacteroides fragilis NCTC 9343
gi|251947250 from Bacteroides sp. 3_2_5
gi|263255266 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **44%**

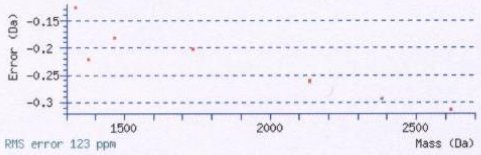
Matched peptides shown in Bold Red

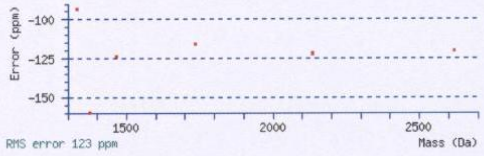
```
  1 MLKTILSISG KPGLYKLISQ GKNMLIVETI DAAKKRFPAY GNEKIISLAD
 51 IAMYTNDSEV PLRDVLRSIK EKENAAIASI DVKKATSEQL REYLAEVLPD
101 FDRDRVYTND IKKLILWYNI LVSNGITDFG EETAVEAE
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
4 - 16	688.8010	1375.5874	1375.8075	-0.2200	0	K.TILSISGKPGLYK.L (Ions score 40)
23 - 34	667.2930	1332.5714	1332.6959	-0.1244	0	K.NMLIVETIDAAK.K Oxidation (M) (Ions score 49)
45 - 63	712.9460	2135.8162	2136.0772	-0.2611	0	K.IISLADIAMYTNDSEVPLR.D Oxidation (M) (Ions score 60)
45 - 63	1068.9160	2135.8174	2136.0772	-0.2598	0	K.IISLADIAMYTNDSEVPLR.D Oxidation (M) (Ions score 102)
45 - 67	874.0220	2619.0442	2619.3578	-0.3136	1	K.IISLADIAMYTNDSEVPLRDVLR.S Oxidation (M) (Ions score 119)
92 - 103	733.7710	1465.5274	1465.7089	-0.1814	0	R.EYLAEVLPDFDR.D (Ions score 56)
92 - 105	579.8860	1736.6362	1736.8369	-0.2008	1	R.EYLAEVLPDFDRDR.V (Ions score 37)






```
LOCUS       YP_098487                138 aa            linear   BCT 26-APR-2009
DEFINITION  hypothetical protein BF1203 [Bacteroides fragilis YCH46].
ACCESSION   YP_098487
VERSION     YP_098487.1  GI:53712495
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 138)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 138)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 138)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD47953.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..138
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..138
```

```
                /product="hypothetical protein"
                /calculated_mol_wt=15396
CDS             1..138
                /locus_tag="BF1203"
                /coded_by="NC_006347.1:1436306..1436722"
                /note="similar to gp:AE016945_239 [Bacteroides
                thetaiotaomicron VPI-5482], percent identity 78 in 138 aa,
                BLASTP E(): 8e-56"
                /transl_table=11
                /db_xref="GeneID:3083031"
```

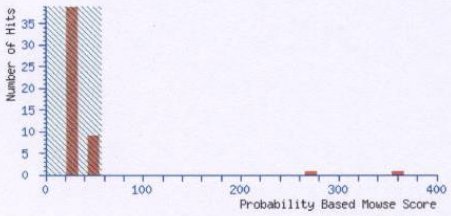
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 5402
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (5402) 10_RF5_01_1076.d\SSP (5402) 10_RF5_01_1076.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:42:08 GMT
Protein hits    : gi|53712495  hypothetical protein BF1203 [Bacteroides fragilis YCH46]
                  gi|255007972 hypothetical protein Bfra3_02458 [Bacteroides fragilis 3_1_12]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53712495 **Mass:** 15517 **Score:** 361 **Queries matched:** 9 **emPAI:** 1.20
hypothetical protein BF1203 [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	103	565.3110	1128.6074	1129.5979	-0.9904	0	30	36	5	K.ENAAIASIDVK.K
☑	120	659.3070	1316.5994	1316.7010	-0.1015	0	(49)	0.35	1	K.NMLIVETIDAAK.K
☑	122	667.2980	1332.5814	1332.6959	-0.1144	0	54	0.11	1	K.NMLIVETIDAAK.K + Oxidation (M)
☑	142	733.8000	1465.5854	1465.7089	-0.1234	0	56	0.067	1	R.EYLAEVLPDFDR.D
☑	173	579.8940	1736.6602	1736.8369	-0.1768	1	43	1.1	1	R.EYLAEVLPDFDRDR.V
☑	220	712.9900	2135.9482	2136.0772	-0.1291	0	(42)	1.3	1	K.IISLADIAMYTNDSEVPLR.D + Oxidation (M)
☑	221	1068.9820	2135.9494	2136.0772	-0.1278	0	86	6.1e-05	1	K.IISLADIAMYTNDSEVPLR.D + Oxidation (M)
☑	222	713.0160	2136.0262	2136.0772	-0.0511	0	(33)	13	1	K.IISLADIAMYTNDSEVPLR.D + Oxidation (M)
☑	247	873.9560	2618.8462	2619.3578	-0.5116	1	93	7.1e-06	1	K.IISLADIAMYTNDSEVPLRDVLR.S + Oxidation (M)

2. gi|255007972 **Mass:** 15498 **Score:** 271 **Queries matched:** 6 **emPAI:** 0.81
hypothetical protein Bfra3_02458 [Bacteroides fragilis 3_1_12]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
142	733.8000	1465.5854	1464.7249	0.8606	0	49	0.29	2	R.QYLAEVLPDFDR.D
173	579.8940	1736.6602	1735.8529	0.8072	1	43	1.1	1	R.QYLAEVLPDFDRDR.V
220	712.9900	2135.9482	2136.0772	-0.1291	0	(42)	1.3	1	K.IISLADIAMYTNDSEVPLR.D + Oxidation (M)
221	1068.9820	2135.9494	2136.0772	-0.1278	0	86	6.1e-05	1	K.IISLADIAMYTNDSEVPLR.D + Oxidation (M)
222	713.0160	2136.0262	2136.0772	-0.0511	0	(33)	13	1	K.IISLADIAMYTNDSEVPLR.D + Oxidation (M)
247	873.9560	2618.8462	2619.3578	-0.5116	1	93	7.1e-06	1	K.IISLADIAMYTNDSEVPLRDVLR.S + Oxidation (M)

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	59	409.7360	817.4574	817.4545	0.0029	0	34	13	1	DVITVSGK
☑	103	565.3110	1128.6074	1128.6437	-0.0363	1	32	20	1	IEAALVRMAR
☑	176	592.5390	1774.5952	1774.9478	-0.3527	1	32	13	1	LYGINASGWTPARLTR
☑	151	521.1690	1560.4852	1559.9623	0.5228	2	31	15	1	VNIILQLAQKHKR
☑	125	677.4280	2029.2622	2029.2339	0.0282	2	30	43	1	KGLTIIPLSIYLKGSWIK
☑	188	944.1040	2829.2902	2830.2758	-0.9856	1	30	38	1	MNTGQWMGTMNLTESGAGSDLGVMKAK + Oxidation (M)
☑	126	678.3130	1354.6114	1355.7159	-1.1044	0	29	39	1	MLFSLFNILDK + Oxidation (M)
	148	763.1340	2286.3802	2286.0872	0.2930	0	28	54	1	MQYDALMDESTTTLSLAIGAR

Mascot Search Results

Protein View

Match to: **gi|53712495** Score: **361**
hypothetical protein BF1203 [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (5402) 10_RF5_01_1076.d\SSP (5402) 10_RF5_01_1076.mgf

Nominal mass (M_r): **15517**; Calculated pI value: **5.15**
NCBI BLAST search of gi|53712495 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60680695 from Bacteroides fragilis NCTC 9343
gi|253563474 from Bacteroides sp. 3_2_5
gi|265762658 from Bacteroides sp. 2_1_16
gi|52215360 from Bacteroides fragilis YCH46
gi|60492129 from Bacteroides fragilis NCTC 9343
gi|251947250 from Bacteroides sp. 3_2_5
gi|263255266 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **43%**

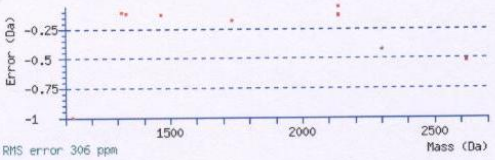
Matched peptides shown in Bold Red

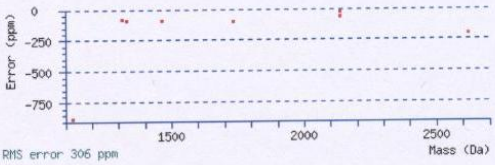
```
  1 MLKTILSISG KPGLYKLISQ GKNMLIVETI DAAKKRFPAY GNEKIISLAD
 51 IAMYTNDSEV PLRDVLRSIK EKENAAIASI DVKKATSEQL REYLAEVLPD
101 FDRDRVYTND IKKLILWYNI LVSNGITDFG EETAVEAE
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
23 - 34	659.3070	1316.5994	1316.7010	-0.1015	0	K.NMLIVETIDAAK.K (Ions score 49)
23 - 34	667.2980	1332.5814	1332.6959	-0.1144	0	K.NMLIVETIDAAK.K Oxidation (M) (Ions score 54)
45 - 63	712.9900	2135.9482	2136.0772	-0.1291	0	K.IISLADIAMYTNDSEVPLR.D Oxidation (M) (Ions score 42)
45 - 63	1068.9820	2135.9494	2136.0772	-0.1278	0	K.IISLADIAMYTNDSEVPLR.D Oxidation (M) (Ions score 86)
45 - 63	713.0160	2136.0262	2136.0772	-0.0511	0	K.IISLADIAMYTNDSEVPLR.D Oxidation (M) (Ions score 33)
45 - 67	873.9560	2618.8462	2619.3578	-0.5116	1	K.IISLADIAMYTNDSEVPLRDVLR.S Oxidation (M) (Ions score 93)
73 - 83	565.3110	1128.6074	1129.5979	-0.9904	0	K.ENAAIASIDVK.K (Ions score 30)
92 - 103	733.8000	1465.5854	1465.7089	-0.1234	0	R.EYLAEVLPDFDR.D (Ions score 56)
92 - 105	579.8940	1736.6602	1736.8369	-0.1768	1	R.EYLAEVLPDFDRDR.V (Ions score 43)






```
LOCUS       YP_098487                138 aa            linear   BCT 26-APR-2009
DEFINITION  hypothetical protein BF1203 [Bacteroides fragilis YCH46].
ACCESSION   YP_098487
VERSION     YP_098487.1  GI:53712495
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 138)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 138)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 138)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD47953.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..138
```

```
                /organism="Bacteroides fragilis YCH46"
                /strain="YCH46"
                /db_xref="taxon:295405"
Protein         1..138
                /product="hypothetical protein"
                /calculated_mol_wt=15396
CDS             1..138
                /locus_tag="BF1203"
                /coded_by="NC_006347.1:1436306..1436722"
                /note="similar to gp:AE016945_239 [Bacteroides
                thetaiotaomicron VPI-5482], percent identity 78 in 138 aa,
                BLASTP E(): 8e-56"
                /transl_table=11
                /db_xref="GeneID:3083031"
```

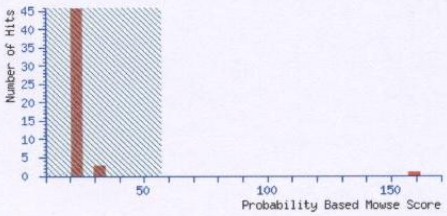
Mascot: http://www.matrixscience.com/

Mascot Search Results

User : Lakshmy Manickan
Email : lakshmy.manickan@unn.ac.uk
Search title : SSP 6102
MS data file : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6102) 10_RF4_01_1074.d\SSP (6102) 10_RF4_01_1074.mgf
Database : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy : Bacteria (Eubacteria) (5690016 sequences)
Timestamp : 6 Jan 2010 at 13:44:14 GMT
Protein hits : gi|53714011 putative RNA-binding protein [Bacteroides fragilis YCH46]

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ Error tolerant

1. gi|53714011 **Mass:** 12893 **Score:** 159 **Queries matched:** 4 **emPAI:** 1.03
putative RNA-binding protein [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	231	813.6560	2437.9462	2438.1060	-0.1598	0	50	0.15	1	K.GFGFVEMPNDEEGNAAIAALNEK.E + Oxidation (M)
☑	244	885.2950	2652.8632	2653.2330	-0.3698	1	70	0.0013	1	R.SKGFGFVEMPNDEEGNAAIAALNEK.E + Oxidation (M)
☑	250	994.3830	2980.1272	2980.3760	-0.2488	1	39	1.8	1	K.GFGFVEMPNDEEGNAAIAALNEKEIDGK.T + Oxidation (M)
	251	994.8610	2981.5612	2980.3760	1.1852	1	(14)	7.5e+02	7	K.GFGFVEMPNDEEGNAAIAALNEKEIDGK.T + Oxidation (M)

Proteins matching the same set of peptides:
gi|255009677 **Mass:** 12909 **Score:** 159 **Queries matched:** 4
putative RNA-binding protein [Bacteroides fragilis 3_1_12]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	131	665.0350	1992.0832	1991.9775	0.1057	1	30	43	1	IYKSSFTAMYTAVTTHR + Oxidation (M)
☑	146	723.3000	2166.8782	2165.9794	0.8988	2	30	52	1	VREVRGGDGGATMDSMDLER + Oxidation (M)
☑	98	370.9340	1109.7802	1110.6761	-0.8959	0	27	46	1	GILAGLGLLER
☑	110	609.5880	1825.7422	1824.9734	0.7688	1	26	1e+02	1	FGAGNDSGPPPQVTLKIK
☑	201	655.2180	1962.6322	1961.9806	0.6516	1	26	43	1	EATGALNDEVLRNLYER
☑	63	385.3540	1153.0402	1153.5074	-0.4672	0	25	98	1	DPSMDALYAR + Oxidation (M)
☑	62	383.3540	1147.0402	1147.6237	-0.5836	0	25	1.2e+02	1	LVLVGDYVDR
☑	180	909.1040	2724.2902	2725.4109	-1.1207	1	24	1.4e+02	1	LDPVAKVAMITNLLTEDNLPSYHR + Oxidation (M)
☑	130	658.3130	1314.6114	1314.6568	-0.0454	0	24	1.1e+02	1	LISATEPDWQR
☑	99	563.1690	1686.4852	1686.7631	-0.2780	1	24	1.4e+02	1	TSGSPSSLFSSSARCR
☑	172	878.1610	2631.4612	2630.3486	1.1126	1	24	1.4e+02	1	IAPDENAGSLHDRLMHLGSDLILK + Oxidation (M)
☑	145	715.1790	1428.3434	1427.7732	0.5702	2	24	1.5e+02	1	ERLNVGKAEDAVK
☑	139	678.9820	2033.9242	2034.0343	-0.1101	1	24	2e+02	1	DIMQELNYELPDLKAVK + Oxidation (M)
☑	165	837.5110	2509.5112	2508.3635	1.1477	2	24	1.8e+02	1	WMQNSITVPLKKLAQVAHSVSR + Oxidation (M)
☑	113	622.6150	1243.2154	1243.6772	-0.4617	0	24	2e+02	1	TDLAALLAATER
☑	95	521.4730	1561.3972	1561.8100	-0.4128	0	23	1.8e+02	1	LAGIAGLGSAAFGSENK
☑	133	666.6390	1996.8952	1996.8943	0.0009	2	23	2e+02	1	KGSDGMGSGGMGSGGKGSDKPK + Oxidation (M)
☑	226	763.6760	2288.0062	2287.0903	0.9158	1	23	97	1	MTSPDRFGGTGSVNIQSGFSVK + Oxidation (M)
☑	108	603.4090	1807.2052	1806.9224	0.2828	1	23	2.2e+02	1	LDGEAIHGKSPNELATR
☑	61	377.1540	752.2934	752.3276	-0.0341	0	23	1e+02	1	AMFGGDR
☑	232	819.3970	2455.1692	2455.1914	-0.0222	1	23	1.1e+02	1	SAPPGEFARELAGHDLAGLSAMSR + Oxidation (M)
	213	1072.7390	3215.1952	3215.7774	-0.5822	1	23	1.5e+02	1	VTLACVGLLIGLGYVLMLAASPAVASRIGASR + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|53714011** Score: **159**
putative RNA-binding protein [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6102) 10_RF4_01_1074.d\SSP (6102) 10_RF4_01_1074.mgf

Nominal mass (M_r): **12893**; Calculated pI value: **5.24**
NCBI BLAST search of gi|53714011 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60682207 from Bacteroides fragilis NCTC 9343
gi|265764361 from Bacteroides sp. 2_1_16
gi|52216876 from Bacteroides fragilis YCH46
gi|60493641 from Bacteroides fragilis NCTC 9343
gi|263256969 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **24%**

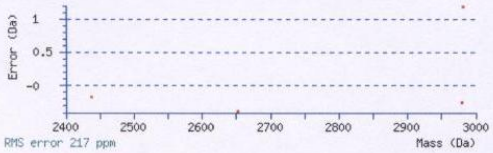
Matched peptides shown in Bold Red

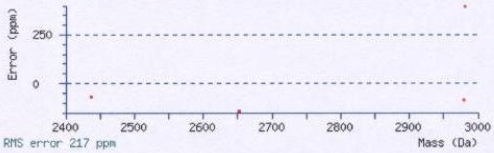
```
  1 MNIFIAGISY NLSNADLGEL FEEFGEVISA KIVMDRETGR SKGFGFVEMP
 51 NDEEGNAAIA ALNEKEIDGK TLAVSVARPR EEGPRRNSNY GGGNRGGYGN
101 NRGGGYGGGN RGGGYGGGRD RY
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
41 - 65	885.2950	2652.8632	2653.2330	-0.3698	1	R.SKGFGFVEMPNDEEGNAAIAALNEK.E Oxidation (M) (Ions score 70)
43 - 65	813.6560	2437.9462	2438.1060	-0.1598	0	K.GFGFVEMPNDEEGNAAIAALNEK.E Oxidation (M) (Ions score 50)
43 - 70	994.3830	2980.1272	2980.3760	-0.2488	1	K.GFGFVEMPNDEEGNAAIAALNEKEIDGK.T Oxidation (M) (Ions score 39)
43 - 70	994.8610	2981.5612	2980.3760	1.1852	1	K.GFGFVEMPNDEEGNAAIAALNEKEIDGK.T Oxidation (M) (Ions score 14)






```
LOCUS       YP_100003                122 aa            linear   BCT 26-APR-2009
DEFINITION  putative RNA-binding protein [Bacteroides fragilis YCH46].
ACCESSION   YP_100003
VERSION     YP_100003.1  GI:53714011
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 122)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 122)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 122)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD49469.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..122
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..122
                     /product="putative RNA-binding protein"
                     /calculated_mol_wt=12770
     Region          2..75
                     /region_name="RRM"
                     /note="RRM (RNA recognition motif), also known as RBD (RNA
                     binding domain) or RNP (ribonucleoprotein domain), is a
                     highly abundant domain in eukaryotes found in proteins
                     involved in post-transcriptional gene expression processes
```

```
                 including mRNA and rRNA...; cd00590"
                 /db_xref="CDD:100104"
Site             order(4,44,46)
                 /site_type="other"
                 /note="RNA/DNA binding site"
                 /db_xref="CDD:100104"
Site             74..75
                 /site_type="other"
                 /note="RRM dimerization site"
                 /db_xref="CDD:100104"
CDS              1..122
                 /locus_tag="BF2719"
                 /coded_by="complement(NC_006347.1:3121075..3121443)"
                 /note="similar to gpu:BX842649_260 [Bdellovibrio
                 bacteriovorus], percent identity 54 in 114 aa, BLASTP E():
                 1e-27"
                 /transl_table=11
                 /db_xref="GeneID:3083599"
```

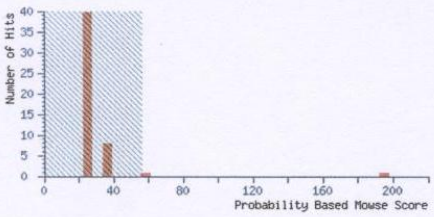
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : LAKSHMY MANICKAN
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 6103
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6103) 10_RH24_01_1296.d\SSP (6103) 10_RH24_01_1296.mgf
Database        : NCBInr 20090522 (8876587 sequences; 3036162093 residues)
Taxonomy        : Bacteria (Eubacteria) (4773688 sequences)
Timestamp       : 30 May 2009 at 10:57:36 GMT
Protein hits    : gi|53714574 lactoylglutathione lyase and related protein [Bacteroides fragilis YCH46]
                  gi|126496   RecName: Full=Lysostaphin; AltName: Full=Glycyl-glycine endopeptidase; Flags: Precursor
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 56 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53714574 **Mass:** 14857 **Score:** 195 **Queries matched:** 3 **emPAI:** 0.85
   lactoylglutathione lyase and related protein [Bacteroides fragilis YCH46]
   ☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	154	498.9330	1493.7772	1493.8354	-0.0582	1	44	1	1	R.KGAEGLNIAFLHPK.S
☑	172	919.4290	1836.8434	1836.9509	-0.1074	0	62	0.015	1	K.SIEEALPYYENVLGLK.C
☑	241	821.6930	2462.0572	2462.2190	-0.1618	0	90	1.7e-05	1	K.GAGVHHVAFAIEDGVANALAEAESK.E

---

2. gi|126496 **Mass:** 51656 **Score:** 64 **Queries matched:** 2 **emPAI:** 0.06
   RecName: Full=Lysostaphin; AltName: Full=Glycyl-glycine endopeptidase; Flags: Precursor
   ☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	161	819.8350	1637.6554	1637.7897	-0.1342	0	37	3.6	1	K.SESASFTPNTDIITR.T
☑	209	1023.0460	2044.0774	2044.9234	-0.8460	0	27	65	1	R.MVNSFSNSTAQDPMPFLK.S + 2 Oxidation (M)

Proteins matching the same set of peptides:
gi|153047 **Mass:** 42213 **Score:** 64 **Queries matched:** 2
lysostaphin (ttg start codon)
gi|3287967 **Mass:** 53058 **Score:** 64 **Queries matched:** 2
RecName: Full=Lysostaphin; AltName: Full=Glycyl-glycine endopeptidase; Flags: Precursor

---

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	164	850.9000	1699.7854	1699.8814	-0.0960	0	42	1.5	1	ANVAEIMNIIGEIEGK
☑	167	858.8630	1715.7114	1715.8777	-0.1662	2	41	1.8	1	ELLNKNHYGMDRVK
☑	175	616.5650	1846.6732	1846.8479	-0.1747	2	30	16	1	MKQEAEANAEADKQER
☑	108	374.4570	1120.3492	1121.5135	-1.1643	0	29	22	1	ASVGMNAAETR + Oxidation (M)
☑	205	679.2120	2034.6142	2034.0204	0.5938	1	29	17	1	LIAEEDAQRVIVFCNTR
☑	184	645.0580	1932.1522	1930.9611	1.1911	0	28	29	1	MWITNGPHADTLVVYAK + Oxidation (M)
☑	93	465.3710	1393.0912	1392.6496	0.4416	1	28	72	1	NEEKYMWIHK + Oxidation (M)
☑	84	394.2260	1179.6562	1179.4835	0.1727	0	28	1e+02	1	LCMMSGGHVR + 2 Oxidation (M)
☑	197	667.5230	1999.5472	1999.9018	-0.3546	1	27	26	1	MAVDGTDHDSPTLDQARR + Oxidation (M)
☑	121	636.6970	1907.0692	1907.9476	-0.8785	0	27	81	1	EVTLSFSEAIAVGDDSIR
☑	213	694.0640	2079.1702	2079.9578	-0.7876	1	26	47	1	NQGMEQRYANEIVGANMR
☑	234	767.3330	2298.9772	2300.1185	-1.1413	1	26	42	1	RYEHYNNTIEHIIDLENK
	248	930.1100	2787.3082	2787.2996	0.0085	1	26	43	1	NMNLAYADVDPRMAVGPEVFSGSGFK + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|53714574** Score: **195**
lactoylglutathione lyase and related protein [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6103) 10_RH24_01_1296.d\SSP (6103) 10_RH24_01_1296.mgf

Nominal mass (M_r): **14857**; Calculated pI value: **5.50**
NCBI BLAST search of gi|53714574 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60682596 from Bacteroides fragilis NCTC 9343
gi|253565078 from Bacteroides sp. 3_2_5
gi|265765962 from Bacteroides sp. 2_1_16
gi|52217439 from Bacteroides fragilis YCH46
gi|60494030 from Bacteroides fragilis NCTC 9343
gi|251946543 from Bacteroides sp. 3_2_5
gi|263253630 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **41%**

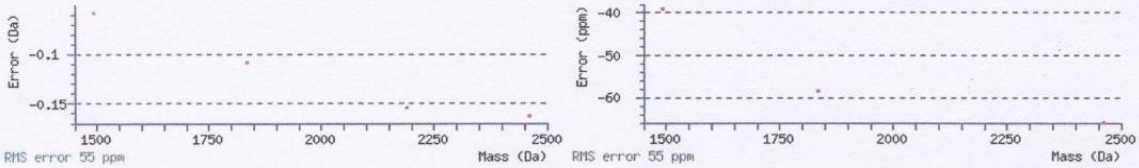
Matched peptides shown in Bold Red

```
  1 MKISHIEHLG IAVKSIEEAL PYYENVLGLK CYNIETVEDQ KVRTAFLKVG
 51 DTKIELLEPT CPESTIAKFI ENKGAGVHHV AFAIEDGVAN ALAEAESKEI
101 RLIDKAPRKG AEGLNIAFLH PKSTLGVLTE LCEH
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
15 - 30	919.4290	1836.8434	1836.9509	-0.1074	0	K.SIEEALPYYENVLGLK.C (Ions score 62)
74 - 98	821.6930	2462.0572	2462.2190	-0.1618	0	K.GAGVHHVAFAIEDGVANALAEAESK.E (Ions score 90)
109 - 122	498.9330	1493.7772	1493.8354	-0.0582	1	R.KGAEGLNIAFLHPK.S (Ions score 44)




```
LOCUS       YP_100566                134 aa            linear   BCT 26-APR-2009
DEFINITION  lactoylglutathione lyase and related protein [Bacteroides fragilis
            YCH46].
ACCESSION   YP_100566
VERSION     YP_100566.1  GI:53714574
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 134)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 134)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 134)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD50032.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
```

```
source          1..134
                /organism="Bacteroides fragilis YCH46"
                /strain="YCH46"
                /db_xref="taxon:295405"
Protein         1..134
                /product="lactoylglutathione lyase and related protein"
                /calculated_mol_wt=14561
Region          5..133
                /region_name="Glyoxalase"
                /note="Glyoxalase/Bleomycin resistance protein/Dioxygenase
                superfamily; cl00411"
                /db_xref="CDD:140684"
CDS             1..134
                /locus_tag="BF3288"
                /coded_by="NC_006347.1:3737763..3738167"
                /note="similar to gp:AE016932_217 [Bacteroides
                thetaiotaomicron VPI-5482], percent identity 98 in 134 aa,
                BLASTP E(): 5e-71"
                /transl_table=11
                /db_xref="GeneID:3082133"
```

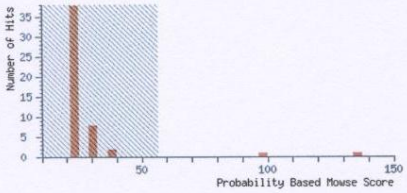
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : LAKSHMY MANICKAN
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 6206
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6206) 10_RJ13_01_1386.d\SSP (6206) 10_RJ13_01_1386.mgf
Database        : NCBInr 20090522 (8876587 sequences; 3036162093 residues)
Taxonomy        : Bacteria (Eubacteria) (4773688 sequences)
Timestamp       : 30 May 2009 at 10:59:43 GMT
Protein hits    : gi|53711592 hypothetical protein BF0301 [Bacteroides fragilis YCH46]
                  gi|143945   Fe-superoxide dismutase [Bacteroides fragilis]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 56 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53711592 **Mass:** 16013 **Score:** 136 **Queries matched:** 2 **emPAI:** 0.47
   hypothetical protein BF0301 [Bacteroides fragilis YCH46]
   ☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	202	1021.4130	2040.8114	2041.0691	-0.2577	1	26	43	1	K.TAPGANDTLTDADALKIVQK.L
☑	234	1139.9630	2277.9114	2277.1198	0.7916	0	112	1.8e-07	1	R.QDLADAELAQVQVMETYLPK.Q + Oxidation (M)

2. gi|143945 **Mass:** 21736 **Score:** 100 **Queries matched:** 3 **emPAI:** 0.15
   Fe-superoxide dismutase [Bacteroides fragilis]
   ☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	239	792.5960	2374.7662	2375.1757	-0.4095	0	27	25	1	K.HLQTYVNNLNSLVPGTEYEGK.T
☑	245	878.6220	2632.8442	2633.3125	-0.4684	0	(25)	35	1	K.LPYANNALEPVISQQTIDYHYGK.H
☑	246	878.6530	2632.9372	2633.3125	-0.3754	0	73	0.00055	1	K.LPYANNALEPVISQQTIDYHYGK.H

Proteins matching the same set of peptides:
gi|533307 **Mass:** 21715 **Score:** 100 **Queries matched:** 3
superoxide dismutase [Bacteroides fragilis]
gi|53713818 **Mass:** 23124 **Score:** 100 **Queries matched:** 3
superoxide dismutase [Bacteroides fragilis YCH46]
gi|60682036 **Mass:** 21758 **Score:** 100 **Queries matched:** 3
superoxide dismutase [Fe] [Bacteroides fragilis NCTC 9343]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	214	700.8960	2099.6662	2100.1262	-0.4601	2	37	2.6	1	YLMVGGDVVDPRVAAQVRR
☑	143	778.5110	2332.5112	2333.1798	-0.6686	1	35	11	1	MIDNIQNVTQDQLRTFIER
☑	235	767.2700	2298.7882	2299.2859	-0.4978	2	31	9.5	1	SGTIKVIGGNLTRNVTITQEAK
☑	153	549.2230	1644.6472	1643.8519	0.7953	0	27	39	1	AFVNPDGTVAEVQVAK
☑	209	694.9150	2081.7232	2080.9338	0.7894	0	27	29	1	TGTGESEEAEEAVGFWQVR
☑	135	742.1990	1482.3834	1481.6425	0.7410	0	26	36	1	TMMGLQNCLNQR + Oxidation (M)
☑	60	396.1590	1185.4552	1184.5509	0.9043	1	26	1.2e+02	1	ANWEGHGKMR
☑	138	757.2680	2268.7822	2268.2114	0.5708	0	26	83	1	KPHLIAILVGNDGGSETYVASK
☑	118	666.3070	1995.8992	1994.9863	0.9128	2	25	1.3e+02	1	AGGFGEAAARRFEWPAFR
	184	940.8210	2819.4412	2820.4004	-0.9592	0	25	93	1	DVIALAASVALSHNTYDAAVFLGVCDK

Mascot Search Results

Protein View

Match to: **gi|53711592** Score: **136**
hypothetical protein BF0301 [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6206) 10_RJ13_01_1386.d\SSP (6206) 10_RJ13_01_1386.mgf

Nominal mass (M_r): **16013**; Calculated pI value: **5.67**
NCBI BLAST search of gi|53711592 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60679842 from Bacteroides fragilis NCTC 9343
gi|253564357 from Bacteroides sp. 3_2_5
gi|265764976 from Bacteroides sp. 2_1_16
gi|52214457 from Bacteroides fragilis YCH46
gi|60491276 from Bacteroides fragilis NCTC 9343
gi|251948133 from Bacteroides sp. 3_2_5
gi|263254360 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **26%**

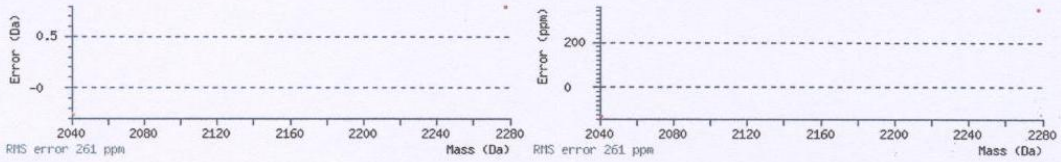
Matched peptides shown in Bold Red

```
  1 MDLFERVSED IKNAMKAKDK VALETLRNVK KFFLEAKTAP GANDTLTDAD
 51 ALKIVQKLVK QGKDAAEIYI GQGRQDLADA ELAQVQVMET YLPKQMSAEE
101 LEAALKEIIA EVGATSGKDM GKVMGVASKK LAGLAEGRAI SAKVKELLG
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
38 - 57	1021.4130	2040.8114	2041.0691	-0.2577	1	K.TAPGANDTLTDADALKIVQK.L (Ions score 26)
75 - 94	1139.9630	2277.9114	2277.1198	0.7916	0	R.QDLADAELAQVQVMETYLPK.Q Oxidation (M) (Ions score 112)




```
LOCUS       YP_097584                149 aa            linear   BCT 26-APR-2009
DEFINITION  hypothetical protein BF0301 [Bacteroides fragilis YCH46].
ACCESSION   YP_097584
VERSION     YP_097584.1  GI:53711592
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 149)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 149)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 149)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD47050.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..149
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..149
                     /product="hypothetical protein"
                     /calculated_mol_wt=15891
```

```
Region          5..148
                /region_name="GatB_Yqey"
                /note="GatB domain; cl11497"
                /db_xref="CDD:142702"
CDS             1..149
                /locus_tag="BF0301"
                /coded_by="complement(NC_006347.1:332834..333283)"
                /note="similar to gp:AE016940_144 [Bacteroides
                thetaiotaomicron VPI-5482], percent identity 91 in 135 aa,
                BLASTP E(): 1e-62"
                /transl_table=11
                /db_xref="GeneID:3082638"
```

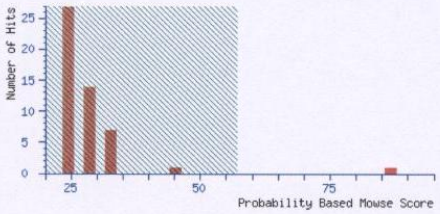
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 6301
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP 6301 (1000)_RN13_01_1592.d\SSP 6301 (1000)_RN13_01_1592.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:46:45 GMT
Protein hits    : gi|53712002 thioredoxin [Bacteroides fragilis YCH46]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< [0.05] Max. number of hits [AUTO]

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off [0] Show sub-sets [0]

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|53712002 **Mass:** 11637 **Score:** 87 **Queries matched:** 2 **emPAI:** 0.30
thioredoxin [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	125	539.7950	1077.5754	1077.6223	-0.0468	0	16	7.2e+02	3	R.NIPTVLFFK.N
☑	184	918.8160	1835.6174	1836.7836 *	-1.1662	0	70	0.0016	1	K.CDVDENSDLPAEFGIR.N

Proteins matching the same set of peptides:
gi|255007522 **Mass:** 11635 **Score:** 85 **Queries matched:** 2
putative thioredoxin [Bacteroides fragilis 3_1_12]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	129	582.2650	1162.5154	1162.6598	-0.1443	0	47	0.6	1	LVNEIITFSK
☑	205	652.9580	1955.8522	1956.9323	-1.0802	0	35	8.2	1	MTDSAGIPGGAGAAGGSSKPPR + Oxidation (M)
☑	126	552.0970	1653.2692	1652.8780	0.3911	2	34	15	1	TSRMLGVSRNTIYR
☑	83	387.2600	1158.7582	1158.5265	0.2317	1	31	52	1	NSDEKSTHNK
☑	114	489.3880	1465.1422	1465.7096	-0.5674	1	29	56	1	GMAARASGQTFDVR
☑	189	934.2050	2799.5932	2799.3940	0.1992	2	28	60	1	DAEALGHRLQQWGAQKADIYSSDLK
☑	134	626.2680	1875.7822	1876.0128	-0.2306	1	27	89	1	VIKINPIQTMTFEQAK + Oxidation (M)
☑	73	366.2200	1095.6382	1094.6448	0.9934	0	27	1.2e+02	1	IILASHGGLSK
☑	137	638.7210	1913.1412	1913.0694	0.0718	1	27	92	1	ALTTQPSLINARLSTTAR
☑	142	652.9630	1303.9114	1303.7347	0.1767	1	26	57	1	TIKQSLSIDTAK
☑	241	1133.4750	3397.4032	3397.6937	-0.2906	1	26	76	1	IEYIEEPYVRASIMVPNDYVGAVMEIAQAK
☑	124	539.2530	1614.7372	1614.9457	-0.2085	2	26	1.5e+02	1	GLYNALKIPKITER
☑	69	354.1940	1059.5602	1060.6063	-1.0461	2	26	1.8e+02	1	AKKNGMIVGK + Oxidation (M)
☑	160	728.9160	2183.7262	2184.2379	-0.5117	2	25	1.3e+02	1	NDGTLRNSALVTKFLLPVAR
☑	164	747.5930	2239.7572	2239.8928	-0.1356	0	25	1.2e+02	1	MAQMFGMTISELCMDTDSR + Oxidation (M)
☑	203	651.2730	1950.7972	1950.0898	0.7074	0	25	73	1	DLARPLGLSVEEAAAAVIR
☑	140	645.2910	1932.8512	1933.1104	-0.2592	2	24	2e+02	1	MKLVIVTGMSGAGKTIALK + Oxidation (M)
☑	198	644.3390	1929.9952	1928.9774	1.0178	1	24	95	1	LPTIRLCMMSVPVPEIG + Oxidation (M)
☑	202	648.5640	1942.6702	1941.9156	0.7546	2	24	72	1	FGFDMGHLGKRVEAHPE + Oxidation (M)
☑	230	1046.2260	3135.6562	3136.5995	-0.9433	2	24	1.5e+02	1	IHPKNPDLVYVAAFGHAFGPNEERGVFR
☑	210	661.0380	1980.0922	1981.0228	-0.9307	1	24	1.1e+02	1	HGASTEDLRAAVTAALAGIAS
☑	170	510.5390	1528.5952	1527.8919	0.7033	2	24	1.1e+02	1	NAKVAAAIVLRGMAK + Oxidation (M)
☑	246	834.3670	2500.0792	2499.2614	0.8178	2	23	86	1	REEMPDMERLSGAGIPVAVFAPK
	148	676.5610	2026.6612	2027.0067	-0.3456	1	23	1.6e+02	1	THDMPIIMVTAKDAEIDK

Mascot Search Results

Protein View

Match to: **gi|53712002** Score: **87**
thioredoxin [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP 6301 (1000)_RN13_01_1592.d\SSP 6301 (1000)_RN13_01_1592.mgf

Nominal mass (M_r): **11637**; Calculated pI value: **4.51**
NCBI BLAST search of gi|53712002 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60680203 from Bacteroides fragilis NCTC 9343
gi|253563963 from Bacteroides sp. 3_2_5
gi|265765340 from Bacteroides sp. 2_1_16
gi|52214867 from Bacteroides fragilis YCH46
gi|60491637 from Bacteroides fragilis NCTC 9343
gi|251947739 from Bacteroides sp. 3_2_5
gi|263254724 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **24%**

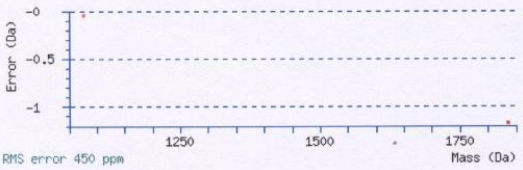
Matched peptides shown in Bold Red

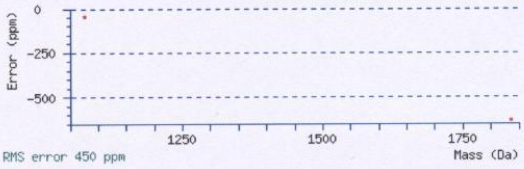
```
  1 MALEITDNNF KEILAEGSPV VIDFWAPWCG PCKMVGPIID ELAKEYEGKV
 51 IMGKCDVDEN SDLPAEFGIR NIPTVLFFKN GELVDKQVGA VGKPAFVEKV
101 EKLL
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
55 - 70	918.8160	1835.6174	1836.7836	-1.1662	0	K.CDVDENSDLPAEFGIR.N (Ions score 70)
71 - 79	539.7950	1077.5754	1077.6223	-0.0468	0	R.NIPTVLFFK.N (Ions score 16)






```
LOCUS       YP_097994                104 aa            linear   BCT 26-APR-2009
DEFINITION  thioredoxin [Bacteroides fragilis YCH46].
ACCESSION   YP_097994
VERSION     YP_097994.1  GI:53712002
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 104)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 104)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 104)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD47460.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..104
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
```

```
Protein         1..104
                /product="thioredoxin"
                /calculated_mol_wt=11339
Region          8..101
                /region_name="TRX_family"
                /note="TRX family; composed of two groups: Group I, which
                includes proteins that exclusively encode a TRX domain;
                and Group II, which are composed of fusion proteins of TRX
                and additional domains. Group I TRX is a small ancient
                protein that alter the redox...; cd02947"
                /db_xref="CDD:48496"
Site            order(29,32)
                /site_type="other"
                /note="catalytic residues"
                /db_xref="CDD:48496"
CDS             1..104
                /locus_tag="BF0713"
                /coded_by="NC_006347.1:838887..839201"
                /note="similar to gp:AE016935_76 [Bacteroides
                thetaiotaomicron VPI-5482], percent identity 81 in 104 aa,
                BLASTP E(): 6e-47"
                /transl_table=11
                /db_xref="GeneID:3082512"
```

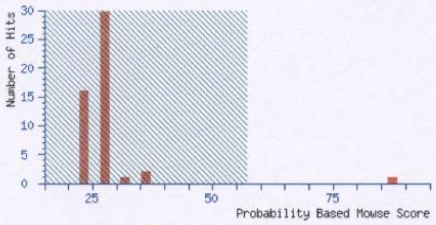
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 6302
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6302) 100_RE17_01_1733.d\SSP (6302) 100_RE17_01_1733.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 16:30:06 GMT
Protein hits    : gi|53715240 triosephosphate isomerase [Bacteroides fragilis YCH46]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53715240 **Mass:** 26867 **Score:** 87 **Queries matched:** 2 **emPAI:** 0.12
triosephosphate isomerase [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	197	547.2600	1638.7582	1638.8478	-0.0896	0	21	2.2e+02	4	K.TASPAQAQEIHAFIR.S
☑	209	928.9290	1855.8434	1855.9680	-0.1245	0	67	0.0056	1	K.ELFANPDVDGGLIGGAALK.V

Proteins matching the same set of peptides:
gi|253566372 **Mass:** 28775 **Score:** 87 **Queries matched:** 2
triosephosphate isomerase [Bacteroides sp. 3_2_5]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	211	623.0900	1866.2482	1865.9781	0.2700	1	35	5.4	1	HNKGIIALSACLAGDVAR
☑	229	687.7600	2060.2582	2060.9875	-0.7293	2	29	27	1	ENNPRNSEGDFITLKDGR
☑	160	654.3460	1960.0162	1960.9524	-0.9362	0	27	1.1e+02	1	AIMTSDLGINPTNDGSAIR + Oxidation (M)
☑	112	438.9630	1313.8672	1312.6888	1.1784	2	27	1.1e+02	1	NRYVFETKTR
☑	182	741.7640	2222.2702	2223.2488	-0.9786	2	26	95	1	LGQKQGHLVTIATGRALFDAK
☑	251	920.2320	2757.6742	2756.5420	1.1322	2	26	34	1	LPLVISLSGIMLLSIGGMWWMRRK
☑	197	547.2600	1638.7582	1639.8603	-1.1021	1	26	63	1	DQKMAPPQVSAEVLK
☑	176	713.3310	2136.9712	2137.0296	-0.0584	1	26	1.3e+02	1	MRYELVDEAGRPMLDQAK + Oxidation (M)
☑	242	763.6450	2287.9132	2288.1908	-0.2777	1	26	46	1	MFLNPMPINSLPTLLGEGGKK + 2 Oxidation (M)
☑	95	365.9130	1094.7172	1095.5084	-0.7912	0	26	1.5e+02	1	AEDLAYDATK
☑	237	718.2730	2151.7972	2151.1647	0.6324	1	26	47	1	AGADIAIERALAEGTSPIAVAR
☑	218	644.7310	1931.1712	1931.9741	-0.8029	2	25	67	1	FKSGYTLGFNRALDSEK
☑	249	878.6190	2632.8352	2633.4615	-0.6263	2	25	47	1	LLFVGDVVGSLGREMVAQYVPKLK + Oxidation (M)
☑	148	618.6470	1852.9192	1853.8982	-0.9790	1	24	1.7e+02	1	EPLMSAIDAFKAFDQR + Oxidation (M)
☑	206	607.2630	1818.7672	1818.9588	-0.1916	2	24	94	1	RDGGAAAKGIGYGELISGK
☑	189	759.3340	2274.9802	2274.9789	0.0013	1	24	1.9e+02	1	MAQGGGMPGMPGMPGMGGGPGRQK + 2 Oxidation (M)
☑	143	561.1460	1680.4162	1681.0515	-0.6353	2	24	1.6e+02	1	HLLLKHNIRTLPVK
☑	139	526.3320	1575.9742	1576.9413	-0.9671	1	24	2.3e+02	1	ASVILLHGLGSRNIK
☑	142	556.5640	1111.1134	1111.5583	-0.4449	0	24	1.8e+02	1	SYPAMTVSLK + Oxidation (M)
☑	234	702.0430	2103.1072	2103.0895	0.0177	2	24	1.1e+02	1	LDMRLFQAAARGGPSGDVVK + Oxidation (M)
☑	225	677.3390	2028.9952	2027.9595	1.0357	2	23	1.1e+02	1	AFFENSQRRMTENNLR + Oxidation (M)
☑	119	477.1930	1428.5572	1427.7633	0.7939	2	23	2.4e+02	1	RNKDGLINQGWK
	220	662.8710	1985.5912	1986.0316	-0.4404	2	23	87	1	RIDLMLEAGALDEARAAR + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|53715240** Score: **87**
triosephosphate isomerase [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6302) 100_RE17_01_1733.d\SSP (6302) 100_RE17_01_1733.mgf

Nominal mass (M_r): **26867**; Calculated pI value: **5.14**
NCBI BLAST search of gi|53715240 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60683174 from Bacteroides fragilis NCTC 9343
gi|81313773 from Bacteroides fragilis NCTC 9343
gi|81824915 from Bacteroides fragilis
gi|52218105 from Bacteroides fragilis YCH46
gi|60494608 from Bacteroides fragilis NCTC 9343

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **13%**

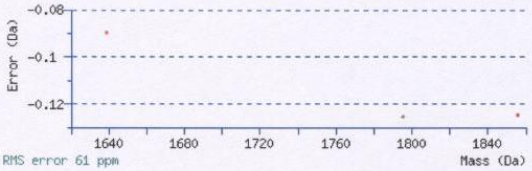
Matched peptides shown in Bold Red

```
  1 MRKNIVAGNW KMNKTLQEGI ALAKELNEAL ANEKPNCDVI ICTPFIHLAS
 51 VTPLVDAAKI GVGAENCADK ESGAYTGEVS AAMVASTGAK YVILGHSERR
101 AYYGETVEIL KDKVKLALAN GLTPIFCIGE VLEEREANKQ NEVVAAQLAS
151 VFDLSAEDFS KIVLAYEPVW AIGTGKTASP AQAQEIHAFI RSAVAEKYGK
201 EIADNTSILY GGSCKPSNAK ELFANPDVDG GLIGGAALKV ADFKGIIDAF
251 N
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
177 - 191	547.2600	1638.7582	1638.8478	-0.0896	0	K.TASPAQAQEIHAFIR.S (Ions score 21)
221 - 239	928.9290	1855.8434	1855.9680	-0.1245	0	K.ELFANPDVDGGLIGGAALK.V (Ions score 67)




```
LOCUS       YP_101232                251 aa            linear   BCT 26-APR-2009
DEFINITION  triosephosphate isomerase [Bacteroides fragilis YCH46].
ACCESSION   YP_101232
VERSION     YP_101232.1  GI:53715240
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 251)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 251)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 251)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD50698.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..251
                     /organism="Bacteroides fragilis YCH46"
```

```
                /strain="YCH46"
                /db_xref="taxon:295405"
Protein         1..251
                /product="triosephosphate isomerase"
                /EC_number="5.3.1.1"
                /calculated_mol_wt=26462
Region          5..248
                /region_name="TIM"
                /note="Triosephosphate isomerase (TIM) is a glycolytic
                enzyme that catalyzes the interconversion of
                dihydroxyacetone phosphate and
                D-glyceraldehyde-3-phosphate. The reaction is very
                efficient and requires neither cofactors nor metal ions.
                TIM, usually...; cd00311"
                /db_xref="CDD:73362"
Site            order(9,11,96,167,173,213,232,234..235)
                /site_type="other"
                /note="substrate binding site"
                /db_xref="CDD:73362"
Site            order(9,12,45..47,49,52,65,83,86..87,98..99)
                /site_type="other"
                /note="dimer interface"
                /db_xref="CDD:73362"
Site            order(11,96,167)
                /site_type="other"
                /note="catalytic triad"
                /db_xref="CDD:73362"
CDS             1..251
                /gene="tpiA"
                /locus_tag="BF3956"
                /coded_by="complement(NC_006347.1:4511582..4512337)"
                /note="Reversibly isomerizes the ketone sugar
                dihydroxyacetone phosphate to the aldehyde sugar
                glyceraldehyde-3-phosphate"
                /transl_table=11
                /db_xref="GeneID:3085037"
```

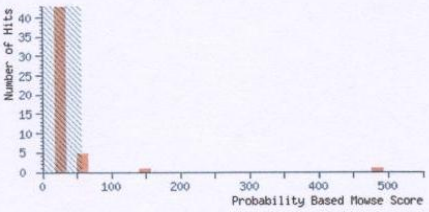
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : LAKSHMY MANICKAN
Email           : lakshmy.manickan@unn.ac.uk
Search title    : 6304 bf
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6302) 10_RJ3_01_1363.d\SSP (6302) 10_RJ3_01_1363.mgf
Database        : NCBInr 20090528 (8915381 sequences; 3049521622 residues)
Taxonomy        : Bacteria (Eubacteria) (4801653 sequences)
Timestamp       : 6 Jun 2009 at 11:57:12 GMT
Protein hits    : gi|53714055  putative thiol peroxidase [Bacteroides fragilis YCH46]
                  gi|160891883 hypothetical protein BACUNI_04341 [Bacteroides uniformis ATCC 8492]
                  gi|167762340 hypothetical protein BACSTE_00694 [Bacteroides stercoris ATCC 43183]
                  gi|218130132 hypothetical protein BACEGG_01719 [Bacteroides eggerthii DSM 20697]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 56 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned Decreasing Score Require bold red ☐

Select All | Select None | Search Selected | ☐ Error tolerant

1. gi|53714055 Mass: 18042 Score: 484 Queries matched: 18 emPAI: 2.93
putative thiol peroxidase [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
104	552.2190	1102.4234	1102.6386	-0.2152	0	76	0.00068	1	K.LIGEFIQVGK.V
117	600.7380	1199.4614	1199.6332	-0.1718	0	68	0.0038	1	R.MADGPLAGLLAR.A + Oxidation (M)
118	601.3060	1200.5974	1199.6332	0.9642	0	(11)	2.4e+03	3	R.MADGPLAGLLAR.A + Oxidation (M)
128	661.1670	1320.3194	1320.5623	-0.2428	0	(16)	7.8e+02	3	R.FSDFDESYGVR.M
129	661.1670	1320.3194	1320.5623	-0.2428	0	(67)	0.0031	1	R.FSDFDESYGVR.M
130	661.1940	1320.3734	1320.5623	-0.1888	0	69	0.0023	1	R.FSDFDESYGVR.M
131	661.2080	1320.4014	1320.5623	-0.1608	0	(23)	86	1	R.FSDFDESYGVR.M
132	661.2180	1320.4214	1320.5623	-0.1408	0	(57)	0.032	1	R.FSDFDESYGVR.M
133	661.2260	1320.4374	1320.5623	-0.1248	0	(58)	0.03	1	R.FSDFDESYGVR.M
184	610.2910	1827.8512	1829.0047	-1.1535	1	50	0.24	1	K.DTVVLAISKDLPFAQGR.F
207	999.3810	1996.7474	1997.9404	-1.1930	0	(34)	5.5	1	R.FCTTEGIENVIPLSDFR.F
209	999.3900	1996.7654	1997.9404	-1.1750	0	(40)	1.6	1	R.FCTTEGIENVIPLSDFR.F
210	999.4050	1996.7954	1997.9404	-1.1450	0	42	0.96	1	R.FCTTEGIENVIPLSDFR.F
226	708.5960	2122.7662	2123.0310	-0.2648	0	37	2.7	1	K.VAYTELVPEITQEPDYEK.A
232	1088.4760	2174.9374	2176.1198	-1.1824	0	(24)	79	3	K.NIVLNIFPSLDTGVCATSVR.K
233	725.9900	2174.9482	2176.1198	-1.1716	0	48	0.29	1	K.NIVLNIFPSLDTGVCATSVR.K
244	808.6340	2422.8802	2423.1744	-0.2942	1	43	0.73	1	K.DGKVAYTELVPEITQEPDYEK.A
245	817.6610	2449.9612	2450.2661	-0.3050	2	54	0.055	1	K.MAAGMKDTVVLAISKDLPFAQGR.F + 2 Oxidation (M)

2. gi|160891883 Mass: 18030 Score: 150 Queries matched: 7 emPAI: 0.41
hypothetical protein BACUNI_04341 [Bacteroides uniformis ATCC 8492]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
117	600.7380	1199.4614	1199.6332	-0.1718	0	68	0.0038	1	R.MADGPLAGLLAR.S + Oxidation (M)
118	601.3060	1200.5974	1199.6332	0.9642	0	(11)	2.4e+03	3	R.MADGPLAGLLAR.S + Oxidation (M)
207	999.3810	1996.7474	1997.9404	-1.1930	0	(34)	5.5	1	R.FCTTEGIENVIPLSDFR.F
209	999.3900	1996.7654	1997.9404	-1.1750	0	(40)	1.6	1	R.FCTTEGIENVIPLSDFR.F
210	999.4050	1996.7954	1997.9404	-1.1450	0	42	0.96	1	R.FCTTEGIENVIPLSDFR.F
232	1088.4760	2174.9374	2176.1198	-1.1824	0	(21)	1.5e+02	7	K.NVILNIFPSLDTGVCATSVR.K
233	725.9900	2174.9482	2176.1198	-1.1716	0	39	2.3	2	K.NVILNIFPSLDTGVCATSVR.K

Mascot Search Results

Protein View

Match to: gi|53714055 Score: 484
putative thiol peroxidase [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6302) 10_RJ3_01_1363.d\SSP (6302) 10_RJ3_01_1363.mgf

Nominal mass (M_r): **18042**; Calculated pI value: **5.55**
NCBI BLAST search of gi|53714055 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60682249 from Bacteroides fragilis NCTC 9343
gi|253564697 from Bacteroides sp. 3_2_5
gi|265766326 from Bacteroides sp. 2_1_16
gi|52216920 from Bacteroides fragilis YCH46
gi|60493683 from Bacteroides fragilis NCTC 9343
gi|251946162 from Bacteroides sp. 3_2_5
gi|263253994 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **68%**

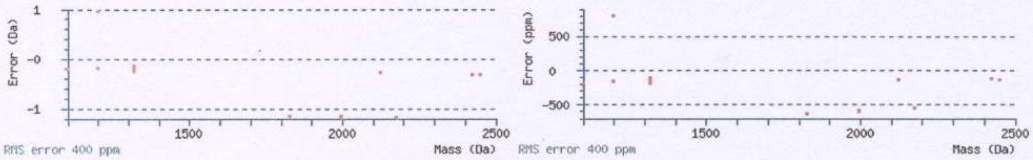
Matched peptides shown in Bold Red

```
  1 MATTNFKGQP VKLIGEFIQV GKVAPDFELV KSDLSSFALK DLKGKNIVLN
 51 IFPSLDTGVC ATSVRKFNKM AAGMKDTVVL AISKDLPFAQ GRFCTTEGIE
101 NVIPLSDFRF SDFDESYGVR MADGPLAGLL ARAVVVIGKD GKVAYTELVP
151 EITQEPDYEK ALAAVK
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
13 - 22	552.2190	1102.4234	1102.6386	-0.2152	0	K.LIGEFIQVGK.V (Ions score 76)
46 - 65	1088.4760	2174.9374	2176.1198	-1.1824	0	K.NIVLNIFPSLDTGVCATSVR.K (Ions score 24)
46 - 65	725.9900	2174.9482	2176.1198	-1.1716	0	K.NIVLNIFPSLDTGVCATSVR.K (Ions score 48)
70 - 92	817.6610	2449.9612	2450.2661	-0.3050	2	K.MAAGMKDTVVLAISKDLPFAQGR.F 2 Oxidation (M) (Ions score 54)
76 - 92	610.2910	1827.8512	1829.0047	-1.1535	1	K.DTVVLAISKDLPFAQGR.F (Ions score 50)
93 - 109	999.3810	1996.7474	1997.9404	-1.1930	0	R.FCTTEGIENVIPLSDFR.F (Ions score 34)
93 - 109	999.3900	1996.7654	1997.9404	-1.1750	0	R.FCTTEGIENVIPLSDFR.F (Ions score 40)
93 - 109	999.4050	1996.7954	1997.9404	-1.1450	0	R.FCTTEGIENVIPLSDFR.F (Ions score 42)
110 - 120	661.1670	1320.3194	1320.5623	-0.2428	0	R.FSDFDESYGVR.M (Ions score 16)
110 - 120	661.1670	1320.3194	1320.5623	-0.2428	0	R.FSDFDESYGVR.M (Ions score 67)
110 - 120	661.1940	1320.3734	1320.5623	-0.1888	0	R.FSDFDESYGVR.M (Ions score 69)
110 - 120	661.2080	1320.4014	1320.5623	-0.1608	0	R.FSDFDESYGVR.M (Ions score 23)
110 - 120	661.2180	1320.4214	1320.5623	-0.1408	0	R.FSDFDESYGVR.M (Ions score 57)
110 - 120	661.2260	1320.4374	1320.5623	-0.1248	0	R.FSDFDESYGVR.M (Ions score 58)
121 - 132	600.7380	1199.4614	1199.6332	-0.1718	0	R.MADGPLAGLLAR.A Oxidation (M) (Ions score 68)
121 - 132	601.3060	1200.5974	1199.6332	0.9642	0	R.MADGPLAGLLAR.A Oxidation (M) (Ions score 11)
140 - 160	808.6340	2422.8802	2423.1744	-0.2942	1	K.DGKVAYTELVPEITQEPDYEK.A (Ions score 43)
143 - 160	708.5960	2122.7662	2123.0310	-0.2648	0	K.VAYTELVPEITQEPDYEK.A (Ions score 37)




```
LOCUS       YP_100047                166 aa            linear   BCT 26-APR-2009
DEFINITION  putative thiol peroxidase [Bacteroides fragilis YCH46].
ACCESSION   YP_100047
VERSION     YP_100047.1  GI:53714055
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 166)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 166)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 166)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD49513.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..166
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..166
                     /product="putative thiol peroxidase"
                     /calculated_mol_wt=17807
     Region          19..164
                     /region_name="PRX_Atyp2cys"
                     /note="Peroxiredoxin (PRX) family, Atypical 2-cys PRX
                     subfamily; composed of PRXs containing peroxidatic and
                     resolving cysteines, similar to the homodimeric thiol
                     specific antioxidant (TSA) protein also known as
                     TRX-dependent thiol peroxidase (Tpx). Tpx is a...;
                     cd03014"
                     /db_xref="CDD:48563"
     Site            order(54,56,85..86,92,108,127..128)
                     /site_type="other"
                     /note="dimer interface"
                     /db_xref="CDD:48563"
     Site            order(57,60,132)
                     /site_type="other"
                     /note="catalytic triad"
                     /db_xref="CDD:48563"
     Site            order(60,94)
                     /site_type="other"
                     /note="peroxidatic and resolving cysteines"
                     /db_xref="CDD:48563"
     CDS             1..166
                     /locus_tag="BF2763"
                     /coded_by="complement(NC_006347.1:3175264..3175764)"
                     /note="similar to gp:AE016931_120 [Bacteroides
                     thetaiotaomicron VPI-5482], percent identity 92 in 166 aa,
                     BLASTP E(): 9e-84"
                     /transl_table=11
                     /db_xref="GeneID:3083562"
```

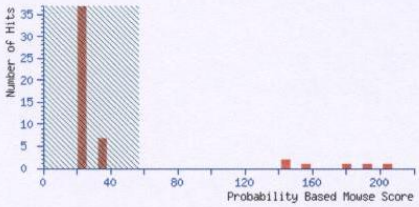
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 6802
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6802) 10_RJ10_01_1377.d\SSP (6802) 10_RJ10_01_1377.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:48:01 GMT
Protein hits    : gi|91215296  3-hydroxybutyryl-CoA dehydrogenase [Psychroflexus torquis ATCC 700755]
                  gi|53714585  fructose-bisphosphate aldolase [Bacteroides fragilis YCH46]
                  gi|255010781 fructose-bisphosphate aldolase [Bacteroides fragilis 3_1_12]
                  gi|29347101  fructose-bisphosphate aldolase [Bacteroides thetaiotaomicron VPI-5482]
                  gi|153808476 hypothetical protein BACCAC_02770 [Bacteroides caccae ATCC 43185]
                  gi|160886425 hypothetical protein BACOVA_04436 [Bacteroides ovatus ATCC 8483]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|91215296 **Mass:** 32980 **Score:** 205 **Queries matched:** 4 **emPAI:** 0.21
3-hydroxybutyryl-CoA dehydrogenase [Psychroflexus torquis ATCC 700755]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 95	393.8730	1178.5972	1178.6845	-0.0874	0	29	33	1	K.AIGMLALPLHK.E + Oxidation (M)
☑ 166	579.9080	1736.7022	1736.8832	-0.1811	0	45	0.64	1	K.DADLLIEAVPENPDVK.T
☑ 189	996.3390	1990.6634	1990.9312	-0.2678	0	79	0.00021	1	K.LGVSNGEGFYTYPEPAYK.S
☑ 229	856.0420	2565.1042	2565.3439	-0.2397	0	51	0.16	1	K.TVFVTNSSTLLPSQFAEVTGRPSK.F

2. gi|53714585 **Mass:** 36766 **Score:** 196 **Queries matched:** 6 **emPAI:** 0.30
fructose-bisphosphate aldolase [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 131	693.3180	1384.6214	1384.7562	-0.1347	1	37	6	1	K.IINVLGSDNKLAE.-
☑ 168	880.8160	1759.6174	1759.8889	-0.2714	0	(17)	3.5e+02	1	R.MVPPPLAFDVLDGVMK.E + 2 Oxidation (M)
☑ 169	587.5700	1759.6882	1759.8889	-0.2007	0	36	6.1	1	R.MVPPPLAFDVLDGVMK.E + 2 Oxidation (M)
☑ 176	919.3450	1836.6754	1837.8516	-1.1762	0	(41)	1.6	1	K.TGCDSLAISIGTSHGAYK.F
☑ 177	613.2370	1836.6892	1837.8516	-1.1625	0	74	0.00079	1	K.TGCDSLAISIGTSHGAYK.F
☑ 222	796.9540	2387.8402	2388.1242	-0.2840	0	49	0.18	1	K.GGYAIPAFNFNNMEQMQAIIK.A + 2 Oxidation (M)

3. gi|255010781 **Mass:** 36694 **Score:** 179 **Queries matched:** 5 **emPAI:** 0.30
fructose-bisphosphate aldolase [Bacteroides fragilis 3_1_12]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
131	693.3180	1384.6214	1384.7562	-0.1347	1	37	6	1	K.IINVLGSDNKLAE.-
169	587.5700	1759.6882	1758.9049	0.7833	0	19	2.9e+02	7	R.MVPPPLAFNVLDGVMK.E + 2 Oxidation (M)
176	919.3450	1836.6754	1837.8516	-1.1762	0	(41)	1.6	1	K.TGCDSLAISIGTSHGAYK.F
177	613.2370	1836.6892	1837.8516	-1.1625	0	74	0.00079	1	K.TGCDSLAISIGTSHGAYK.F
222	796.9540	2387.8402	2388.1242	-0.2840	0	49	0.18	1	K.GGYAIPAFNFNNMEQMQAIIK.A + 2 Oxidation (M)

4. gi|29347101 **Mass:** 36791 **Score:** 160 **Queries matched:** 4 **emPAI:** 0.30
fructose-bisphosphate aldolase [Bacteroides thetaiotaomicron VPI-5482]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
131	693.3180	1384.6214	1384.7562	-0.1347	1	37	6	1	K.ILNVLGSDNKLAE.-
176	919.3450	1836.6754	1837.8516	-1.1762	0	(41)	1.6	1	R.TGCDSLAISIGTSHGAYK.F
177	613.2370	1836.6892	1837.8516	-1.1625	0	74	0.00079	1	R.TGCDSLAISIGTSHGAYK.F
222	796.9540	2387.8402	2388.1242	-0.2840	0	49	0.18	1	K.GGYAIPAFNFNNMEQMQAIIK.A + 2 Oxidation (M)

---

5. gi|153808476 **Mass:** 36778 **Score:** 148 **Queries matched:** 4 **emPAI:** 0.30
hypothetical protein BACCAC_02770 [Bacteroides caccae ATCC 43185]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
131	693.3180	1384.6214	1383.7722	0.8493	1	25	1e+02	3	K.ILNVLGSDNKLAQ.-
176	919.3450	1836.6754	1837.8516	-1.1762	0	(41)	1.6	1	R.TGCDSLAISIGTSHGAYK.F
177	613.2370	1836.6892	1837.8516	-1.1625	0	74	0.00079	1	R.TGCDSLAISIGTSHGAYK.F
222	796.9540	2387.8402	2388.1242	-0.2840	0	49	0.18	1	K.GGYAIPAFNFNNMEQMQAIIK.A + 2 Oxidation (M)

Proteins matching the same set of peptides:
gi|255693155 **Mass:** 36831 **Score:** 148 **Queries matched:** 4
fructose-1,6-bisphosphate aldolase, class II [Bacteroides finegoldii DSM 17565]
gi|260173446 **Mass:** 36803 **Score:** 148 **Queries matched:** 4
fructose-bisphosphate aldolase [Bacteroides sp. D2]

---

6. gi|160886425 **Mass:** 36875 **Score:** 148 **Queries matched:** 4 **emPAI:** 0.29
hypothetical protein BACOVA_04436 [Bacteroides ovatus ATCC 8483]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
131	693.3180	1384.6214	1383.7722	0.8493	1	25	1e+02	3	K.IINVLGSDNKLAQ.-
176	919.3450	1836.6754	1837.8516	-1.1762	0	(41)	1.6	1	R.TGCDSLAISIGTSHGAYK.F
177	613.2370	1836.6892	1837.8516	-1.1625	0	74	0.00079	1	R.TGCDSLAISIGTSHGAYK.F
222	796.9540	2387.8402	2388.1242	-0.2840	0	49	0.18	1	K.GGYAIPAFNFNNMEQMQAIIK.A + 2 Oxidation (M)

Proteins matching the same set of peptides:
gi|237713705 **Mass:** 36861 **Score:** 148 **Queries matched:** 4
fructose-bisphosphate aldolase [Bacteroides sp. D1]
gi|237723289 **Mass:** 36773 **Score:** 148 **Queries matched:** 4
fructose-bisphosphate aldolase [Bacteroides sp. 2_2_4]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	144	755.2960	2262.8662	2262.1541	0.7121	2	33	20	1	SEALFRNMMTGFKYTLPIK + Oxidation (M)
☑	170	895.1500	2682.4282	2683.4302	-1.0020	2	29	50	1	RDGMLIPYMTLSPAIHQGIVARTK + Oxidation (M)
☑	197	1033.8310	2065.6474	2066.0643	-0.4169	2	27	63	1	AGAEAKPEAKSEAKSAPAPEK
☑	91	584.8950	1751.6632	1751.8045	-0.1413	0	27	85	1	MDTGMVYINHPTMVK + Oxidation (M)
☑	147	766.4820	2296.4242	2295.2773	1.1469	2	27	98	1	PLPKRALLAGFSNVLEQAMPK + Oxidation (M)
☑	210	735.6100	2203.8082	2202.9807	0.8275	0	26	44	1	VMSTDLSNGMMATTVQGSDVK + 2 Oxidation (M)
☑	88	563.1570	1686.4492	1685.7984	0.6508	0	25	1.1e+02	1	AAYGMPPHTWLTDAR
☑	116	656.7810	1967.3212	1966.8587	0.4625	0	25	1.3e+02	1	MPADHMVLMAGFTAGNEK + 3 Oxidation (M)
☑	105	634.6000	1900.7782	1901.9694	-1.1913	0	25	1.4e+02	1	IDLTGSSSGNGTLSDPLLR
☑	96	592.6870	1775.0392	1774.0427	0.9965	0	24	1.9e+02	1	ILIVGLGVIGGSFAMALK + Oxidation (M)
☑	175	915.6430	2743.9072	2743.3534	0.5538	2	24	1.3e+02	1	DMPHRLQFTGPTGTNAVEAAMKLAR + 2 Oxidation (M)
☑	58	406.0880	1215.2422	1214.6003	0.6418	1	24	1.5e+02	1	GSPQAGGSEARAK
☑	64	831.2410	830.2337	830.4610	-0.2273	1	24	1.4e+02	1	VRGSGDIK
☑	167	582.5710	1744.6912	1744.8903	-0.1991	2	24	96	1	RSNAILAMRNSYHGR
☑	143	751.7510	2252.2312	2251.0475	1.1836	1	24	1.9e+02	1	MLFNGLCLFEQASLCFRK + Oxidation (M)
☑	70	453.7260	905.4374	905.4463	-0.0088	0	24	1.5e+02	1	DLMLAGMR
☑	109	640.6990	1919.0752	1918.9232	0.1520	1	24	2e+02	1	KAVADVAGESSISDVSGDGR
☑	152	798.8320	2393.4742	2393.2638	0.2104	2	23	1.9e+02	1	AMLLKPTKDDVQGLGFHSHKR + Oxidation (M)
☑	165	865.3170	2592.9292	2592.3193	0.6099	2	23	1.6e+02	1	ANMTFHKGKLYGQDVVAVVCGIGK
☑	223	800.5750	2398.7032	2398.2566	0.4466	1	23	77	1	EYGELMRAAAGGGLITVATVYLK + Oxidation (M)
☑	87	563.0610	1686.1612	1685.6228	0.5384	0	23	1.9e+02	1	APAAPAMDPSMGMGGMM + 4 Oxidation (M)
☑	212	1126.9670	2251.9194	2251.1267	0.7928	1	23	96	1	QAFSDLTATLLDNLGMETRR
☑	112	647.4630	1939.3672	1939.8218	-0.4546	0	23	2e+02	1	MSEFGDSASAGTAGRPTSPD
☑	137	479.2240	1434.6502	1435.7558	-1.1057	0	23	1.5e+02	1	LTDADVGALASYLK
☑	150	780.5170	2338.5292	2339.1411	-0.6119	2	23	2.2e+02	1	YRVQMMFAVPTMYRMLVR + 3 Oxidation (M)
☑	75	498.2930	994.5714	994.4906	0.0808	0	22	1.5e+02	1	MGIFAAAGGGK + Oxidation (M)
☑	163	858.1400	2571.3982	2571.2751	0.1230	1	22	2.3e+02	1	VRAAFEDQDAMVVVHDAIVTNQK + Oxidation (M)
☑	191	996.9930	2987.9572	2988.4970	-0.5398	1	22	2.2e+02	1	SLHTIKSDVLNLTDDTMPSMMLVNGIK + Oxidation (M)
☑	173	912.2700	2733.7882	2733.3771	0.4111	0	22	1.8e+02	1	VMGGLLATGIASAPFLWSMLHDYQR
☑	190	664.5630	1990.6672	1989.9353	0.7318	1	22	1e+02	1	EEEVPDWKQMQDSILK + Oxidation (M)
☑	139	731.8500	2192.5282	2193.2191	-0.6909	2	22	2.8e+02	1	MPDAKPVVPKDGAKITLVNGK + Oxidation (M)
☑	135	475.2010	1422.5812	1423.6806	-1.0994	1	22	1.6e+02	1	YFGMNKGITSYK + Oxidation (M)
☑	86	562.7810	1123.5474	1122.7125	0.8350	1	22	1.7e+02	1	IVLQAQKVPK
☑	181	942.9120	2825.7142	2826.3705	-0.6563	2	22	2.4e+02	1	NSYDMAETSRDTAVELLKEILEQR + Oxidation (M)
☑	203	710.8680	2129.5822	2129.2321	0.3501	1	22	1.1e+02	1	PEPRILVLGAGPAGTATALGLR
☑	208	1091.6490	3271.9252	3271.6673	0.2579	1	22	2.3e+02	1	QVPGLPAQPGAGGGHMSPALVAQAWNRYAGGLK + Oxidation (M)
☑	184	641.3350	1920.9832	1921.9720	-0.9888	1	22	1.8e+02	1	RAGYATVVAMLDLDHFK + Oxidation (M)
☑	76	499.2150	996.4154	995.4971	0.9184	2	22	1.7e+02	1	RMKAFGDR + Oxidation (M)
	193	1009.2650	3024.7732	3024.4507	0.3224	2	21	2.2e+02	1	KLFGSQATVVTSGHAWLDMMPMSADKGK + 2 Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|53714585** Score: **196**
fructose-bisphosphate aldolase [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6802) 10_RJ10_01_1377.d\SSP (6802) 10_RJ10_01_1377.mgf

Nominal mass (M_r): **36766**; Calculated pI value: **5.37**
NCBI BLAST search of gi|53714585 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60682607 from Bacteroides fragilis NCTC 9343
gi|253565089 from Bacteroides sp. 3_2_5
gi|265765951 from Bacteroides sp. 2_1_16
gi|52217450 from Bacteroides fragilis YCH46
gi|60494041 from Bacteroides fragilis NCTC 9343
gi|251946554 from Bacteroides sp. 3_2_5
gi|263253619 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **20%**

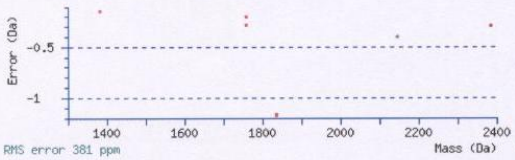
Matched peptides shown in Bold Red

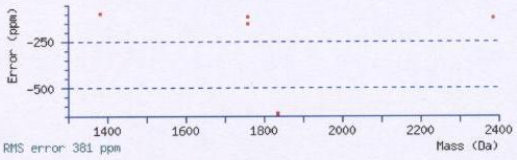
```
  1 MVNYKDLGLV NTRDMFAKAI KGGYAIPAFN FNNMEQMQAI IKAAVETKSP
 51 VILQVSKGAR QYANATLLRY MAQGAVEYAK ELGCKNPEIV LHLDHGDTFE
101 TCKSCIDSGF SSVMIDGSHL PYDENVALTK KVVEYAHQFD VTVEGELGVL
151 AGVEDEVSSD HHTYTEPDEV VDFVTKTGCD SLAISIGTSH GAYKFTPEQC
201 HIDPKTGRMV PPPLAFDVLD GVMKELPGFP IVLHGSSSVP EEEVATINQF
251 GGALKAAIGI PEEELRKAAK SAVCKINIDS DSRLAMTAAI RKVFAEKPAE
301 FDPRKYLGPA RDNMEKLYKH KIINVLGSDN KLAE
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
22 - 42	796.9540	2387.8402	2388.1242	-0.2840	0	K.GGYAIPAFNFNNMEQMQAIIK.A 2 Oxidation (M) (Ions score 49)
177 - 194	919.3450	1836.6754	1837.8516	-1.1762	0	K.TGCDSLAISIGTSHGAYK.F (Ions score 41)
177 - 194	613.2370	1836.6892	1837.8516	-1.1625	0	K.TGCDSLAISIGTSHGAYK.F (Ions score 74)
209 - 224	880.8160	1759.6174	1759.8889	-0.2714	0	R.MVPPPLAFDVLDGVMK.E 2 Oxidation (M) (Ions score 17)
209 - 224	587.5700	1759.6882	1759.8889	-0.2007	0	R.MVPPPLAFDVLDGVMK.E 2 Oxidation (M) (Ions score 36)
322 - 334	693.3180	1384.6214	1384.7562	-0.1347	1	K.IINVLGSDNKLAE.- (Ions score 37)






```
LOCUS       YP_100577                334 aa            linear   BCT 26-APR-2009
DEFINITION  fructose-bisphosphate aldolase [Bacteroides fragilis YCH46].
ACCESSION   YP_100577
VERSION     YP_100577.1  GI:53714585
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 334)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 334)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 334)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD50043.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
```

```
source          1..334
                /organism="Bacteroides fragilis YCH46"
                /strain="YCH46"
                /db_xref="taxon:295405"
Protein         1..334
                /product="fructose-bisphosphate aldolase"
                /EC_number="4.1.2.13"
                /calculated_mol_wt=36311
Region          12..330
                /region_name="TBP_aldolase_IIB"
                /note="Tagatose-1,6-bisphosphate (TBP) aldolase and
                related Type B Class II aldolases. TBP aldolase is a
                tetrameric class II aldolase that catalyzes the reversible
                condensation of dihydroxyacetone phosphate with
                glyceraldehyde 3-phsophate to produce tagatose...;
                cd00947"
                /db_xref="CDD:29573"
Site            order(33..35,38,58,61..63,67,71,74..75,154,280..281,
                283..284,287,291,294)
                /site_type="other"
                /note="intersubunit interface"
                /db_xref="CDD:29573"
Site            order(94..95,189..191,193,234..235,237,277,279..280)
                /site_type="active"
                /db_xref="CDD:29573"
Site            order(95,190,234)
                /site_type="other"
                /note="zinc binding site"
                /db_xref="CDD:29573"
Site            order(189,191,193,235,237)
                /site_type="other"
                /note="Na+ binding site"
                /db_xref="CDD:29573"
CDS             1..334
                /locus_tag="BF3299"
                /coded_by="complement(NC_006347.1:3750977..3751981)"
                /note="catalyzes the formation of glycerone phosphate and
                glyceraldehyde 3-phosphate from fructose 1,6,
                bisphosphate"
                /transl_table=11
                /db_xref="GeneID:3082144"
```

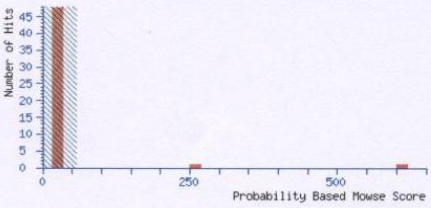
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 6807
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6807) 10_RJ8_01_1373.d\SSP (6807) 10_RJ8_01_1373.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 13:49:18 GMT
Protein hits    : gi|53712366  hypothetical protein BF1074 [Bacteroides fragilis YCH46]
                  gi|255007873 hypothetical protein Bfra3_01963 [Bacteroides fragilis 3_1_12]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53712366 **Mass:** 49095 **Score:** 610 **Queries matched:** 14 **emPAI:** 0.68
hypothetical protein BF1074 [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	88	501.7010	1001.3874	1001.4892	-0.1017	0	38	4.8	1	K.DLFMLGYK.A + Oxidation (M)
☑	90	547.7660	1093.5174	1093.5808	-0.0633	0	70	0.0026	1	K.AVDYIAFAPK.V
☑	108	675.7900	1349.5654	1349.6616	-0.0961	0	55	0.082	1	K.VDVNQAYDWLK.K
	121	**493.5310**	**1477.5712**	**1477.7565**	**-0.1853**	**1**	**(20)**	**2.2e+02**	**4**	**K.VDVNQAYDWLKK.S**
☑	122	739.8240	1477.6334	1477.7565	-0.1231	1	21	2.4e+02	1	K.VDVNQAYDWLKK.S
☑	123	752.3240	1502.6334	1502.8093	-0.1758	0	58	0.043	1	R.ITIGGWIGESTTIR.-
	130	**518.3990**	**1552.1752**	**1552.6980**	**-0.5228**	**0**	**12**	**1.1e+03**	**10**	**K.YFDELMNTHDLR.M**
☑	137	813.7950	1625.5754	1625.7573	-0.1818	0	110	2e-07	1	K.FFDEAINLETDNAK.K
☑	185	973.3220	1944.6294	1945.8040	-1.1746	0	74	0.00066	1	K.CTDSDAYQQASFYVYK.I
☑	102	649.2260	1944.6562	1945.8040	-1.1478	0	(24)	1.7e+02	1	K.CTDSDAYQQASFYVYK.I
☑	205	682.6080	2044.8022	2044.9564	-0.1542	0	(30)	18	1	K.AESAAATLFYFLQMSHDK.L + Oxidation (M)
☑	206	1023.4270	2044.8394	2044.9564	-0.1170	0	61	0.015	1	K.AESAAATLFYFLQMSHDK.L + Oxidation (M)
☑	244	894.9800	2681.9182	2682.2847	-0.3665	0	57	0.029	1	K.DNLVALFINSGTADCESLQGIYGPK.V
☑	251	987.7030	2960.0872	2960.3927	-0.3055	1	55	0.04	1	K.EQFIQDYLAASEYADDAIAAADKESVK.K

Proteins matching the same set of peptides:
gi|60491969 **Mass:** 48413 **Score:** 610 **Queries matched:** 14
conserved hypothetical protein [Bacteroides fragilis NCTC 9343]

---

2. gi|255007873 **Mass:** 49151 **Score:** 267 **Queries matched:** 4 **emPAI:** 0.21
hypothetical protein Bfra3_01963 [Bacteroides fragilis 3_1_12]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	88	501.7010	1001.3874	1001.4892	-0.1017	0	38	4.8	1	K.DLFMLGYK.A + Oxidation (M)
	90	547.7660	1093.5174	1093.5808	-0.0633	0	70	0.0026	1	K.AVDYIAFAPK.M
	137	813.7950	1625.5754	1625.7573	-0.1818	0	110	2e-07	1	K.FFDEAINLETDNAK.K
☑	189	981.2920	1960.5694	1959.8196	0.7498	0	49	0.23	1	K.CTDSEAYQQASFYVYK.I

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	232	766.9760	2297.9062	2297.1321	0.7740	1	33	8.4	1	SGMIVSDNGTEFTSNAILARSK
☑	91	555.8140	1109.6134	1108.5699	1.0435	0	33	13	1	IGVGHPSMAPK + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|60491969** Score: **610**
conserved hypothetical protein [Bacteroides fragilis NCTC 9343]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6807) 10_RJ8_01_1373.d\SSP (6807) 10_RJ8_01_1373.mgf

Nominal mass (M_r): **48413**; Calculated pI value: **5.79**
NCBI BLAST search of gi|60491969 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis NCTC 9343

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **35%**

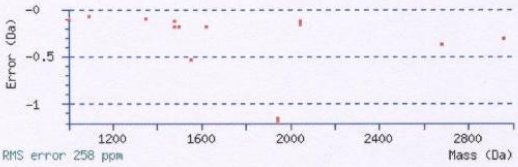
Matched peptides shown in Bold Red

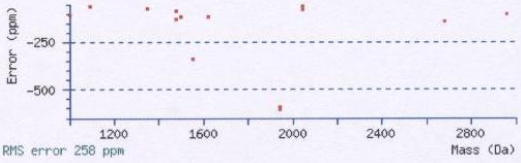
```
  1 MAVLFLSAGA TTVVAQDDAN CNSNSSISHE AVKAGNFKDA YTPWKAVLEN
 51 CPTLRFYTFT DGYKILKGLL GQIKDRNSAE YKKYFDELMN THDLRMKYTQ
101 EFLGKGVKVS SEDEALGIKA VDYIAFAPKV DVNQAYDWLK KSVDAAKAES
151 AAATLFYFLQ MSHDKLKEDP AHKEQFIQDY LAASEYADDA IAAADKESVK
201 KAFGGIKDNL VALFINSGTA DCESLQGIYG PKVETNQTDL NYLKKVISIM
251 KMMKCTDSDA YQQASFYVYK IEPSAEAATG CAYQAYKKGD IDGSVKFFDE
301 AINLETDNAK KAEKAYAAAS VLTTAKKLSQ ARSYAQKAIS FNENYGAPYI
351 LIANLYAMSP NWSDESALNK CTYFAVIDKL QKAKSVDPSV TEEVNKMISR
401 YSAYTPQAKD LFMLGYKAGD RITIGGWIGE STTIR
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
84 - 95	518.3990	1552.1752	1552.6980	-0.5228	0	K.YFDELMNTHDLR.M (Ions score 12)
120 - 129	547.7660	1093.5174	1093.5808	-0.0633	0	K.AVDYIAFAPK.V (Ions score 70)
130 - 140	675.7900	1349.5654	1349.6616	-0.0961	0	K.VDVNQAYDWLK.K (Ions score 55)
130 - 141	493.5310	1477.5712	1477.7565	-0.1853	1	K.VDVNQAYDWLKK.S (Ions score 20)
130 - 141	739.8240	1477.6334	1477.7565	-0.1231	1	K.VDVNQAYDWLKK.S (Ions score 21)
148 - 165	682.6080	2044.8022	2044.9564	-0.1542	0	K.AESAAATLFYFLQMSHDK.L Oxidation (M) (Ions score 30)
148 - 165	1023.4270	2044.8394	2044.9564	-0.1170	0	K.AESAAATLFYFLQMSHDK.L Oxidation (M) (Ions score 61)
174 - 200	987.7030	2960.0872	2960.3927	-0.3055	1	K.EQFIQDYLAASEYADDAIAAADKESVK.K (Ions score 55)
208 - 232	894.9800	2681.9182	2682.2847	-0.3665	0	K.DNLVALFINSGTADCESLQGIYGPK.V (Ions score 57)
255 - 270	973.3220	1944.6294	1945.8040	-1.1746	0	K.CTDSDAYQQASFYVYK.I (Ions score 74)
255 - 270	649.2260	1944.6562	1945.8040	-1.1478	0	K.CTDSDAYQQASFYVYK.I (Ions score 24)
297 - 310	813.7950	1625.5754	1625.7573	-0.1818	0	K.FFDEAINLETDNAK.K (Ions score 110)
410 - 417	501.7010	1001.3874	1001.4892	-0.1017	0	K.DLFMLGYK.A Oxidation (M) (Ions score 38)
422 - 435	752.3240	1502.6334	1502.8093	-0.1758	0	R.ITIGGWIGESTTIR.- (Ions score 58)






```
LOCUS       CAH06730                 435 aa            linear   BCT 13-MAY-2009
DEFINITION  conserved hypothetical protein [Bacteroides fragilis NCTC 9343].
ACCESSION   CAH06730
VERSION     CAH06730.1  GI:60491969
DBSOURCE    embl accession CR626927.1
KEYWORDS    .
SOURCE      Bacteroides fragilis NCTC 9343
  ORGANISM  Bacteroides fragilis NCTC 9343
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 435)
  AUTHORS   Cerdeno-Tarraga,A.M., Patrick,S., Crossman,L.C., Blakely,G.,
            Abratt,V., Lennard,N., Poxton,I., Duerden,B., Harris,B.,
            Quail,M.A., Barron,A., Clark,L., Corton,C., Doggett,J.,
            Holden,M.T., Larke,N., Line,A., Lord,A., Norbertczak,H., Ormond,D.,
            Price,C., Rabbinowitsch,E., Woodward,J., Barrell,B. and Parkhill,J.
  TITLE     Extensive DNA inversions in the B. fragilis genome control variable
            gene expression
  JOURNAL   Science 307 (5714), 1463-1465 (2005)
   PUBMED   15746427
REFERENCE   2  (residues 1 to 435)
  AUTHORS   Cerdeno-Tarraga,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2004) Cerdeno-Tarraga A.M., submitted on behalf
            of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust
            Genome Campus, Hinxton, Cambridge CB10 1SA E-mail:
```

```
            amct@sanger.ac.uk
FEATURES             Location/Qualifiers
     source          1..435
                     /organism="Bacteroides fragilis NCTC 9343"
                     /strain="ATCC 25285 = NCTC 9343"
                     /db_xref="taxon:272559"
     Protein         1..435
                     /product="conserved hypothetical protein"
     CDS             1..435
                     /locus_tag="BF0991"
                     /coded_by="CR626927.1:1241398..1242705"
                     /note="Similar to Bacteroides thetaiotaomicron
                     hypothetical protein BT4368 SWALL:Q89ZK9 (EMBL:AE016945)
                     (433 aa) fasta scores: E(): 8.7e-145, 85.21% id in 433 aa"
                     /transl_table=11
                     /db_xref="GOA:Q5LGK9"
                     /db_xref="InterPro:IPR013026"
                     /db_xref="UniProtKB/TrEMBL:Q5LGK9"
```

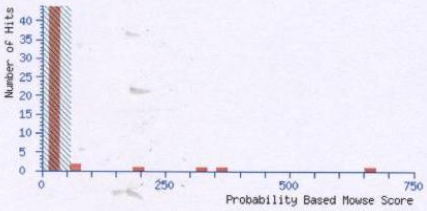
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User           : Lakshmy Manickan
Email          : lakshmy.manickan@unn.ac.uk
Search title   : SSP 6901
MS data file   : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6901) 10_RH23_01_1294.d\SSP (6901) 10_RH23_01_1294.mgf
Database       : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy       : Bacteria (Eubacteria) (5690016 sequences)
Timestamp      : 7 Jan 2010 at 16:32:37 GMT
Protein hits   : gi|53715725  phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]
                 gi|150003549 phosphoenolpyruvate carboxykinase [Bacteroides vulgatus ATCC 8482]
                 gi|150009397 phosphoenolpyruvate carboxykinase [Parabacteroides distasonis ATCC 8503]
                 gi|34541315  phosphoenolpyruvate carboxykinase [Porphyromonas gingivalis W83]
                 gi|123444162 phosphoenolpyruvate carboxykinase [Yersinia enterocolitica subsp. enterocolitica 8081]
                 gi|281421519 phosphoenolpyruvate carboxykinase [Prevotella copri DSM 18205]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|53715725 **Mass:** 59339 **Score:** 663 **Queries matched:** 14 **emPAI:** 0.71
phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	115	684.2770	1366.5394	1365.6677	0.8717	0	24	89	3	K.AYLVNTGWNGSGK.R
☑	120	708.3170	1414.6194	1414.7304	-0.1109	0	93	1.4e-05	1	R.DALLENVTVAADGK.I
☑	141	811.3150	1620.6154	1620.7098	-0.0944	0	47	0.4	1	K.GMFSIMNYMNPLR.G + 3 Oxidation (M)
☑	142	818.3880	1634.7614	1634.8515	-0.0901	0	110	2.5e-07	1	K.ELGLNSETATVFNLK.T
☑	151	843.3600	1684.7054	1685.7719	-1.0665	0	32	15	1	R.LFVVDTFCGANEGTR.M
☑	153	863.9200	1725.8254	1725.9512	-0.1258	1	60	0.029	1	R.GIIDAILDGSIDKAPTK.V
☑	186	972.9460	1943.8774	1944.0204	-0.1429	1	53	0.14	1	K.VINLDKESEPDIFNAIK.R
☑	190	656.6100	1966.8082	1966.9783	-0.1701	0	(46)	0.52	1	K.GQVTELGAVNVMTGVYTGR.S + Oxidation (M)
☑	191	984.4190	1966.8234	1966.9783	-0.1548	0	115	6.2e-08	1	K.GQVTELGAVNVMTGVYTGR.S + Oxidation (M)
☑	195	993.4470	1984.8794	1985.0986	-0.2191	0	(59)	0.05	1	R.VSYPIYHIENIVKPVSK.G
☑	196	662.6660	1984.9762	1985.0986	-0.1224	0	62	0.015	1	R.VSYPIYHIENIVKPVSK.G
	215	705.9670	2114.8792	2115.1293	-0.2501	0	(16)	5.5e+02	5	K.VIPFFDFVVPTELPGVDPK.I
☑	216	1058.5150	2115.0154	2115.1293	-0.1138	0	41	1.9	1	K.VIPFFDFVVPTELPGVDPK.I
☑	250	989.3890	2965.1452	2966.2732	-1.1281	0	33	7.2	1	R.GIASMHCSANTDMEGTSSAIFFGLSGTGK.T + 2 Oxidation (M)

2. gi|150003549 **Mass:** 59400 **Score:** 366 **Queries matched:** 6 **emPAI:** 0.31
phosphoenolpyruvate carboxykinase [Bacteroides vulgatus ATCC 8482]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
141	811.3150	1620.6154	1620.7098	-0.0944	0	47	0.4	1	K.GMFSIMNYMNPLR.G + 3 Oxidation (M)
142	818.3880	1634.7614	1634.8515	-0.0901	0	110	2.5e-07	1	K.ELGLNSETATVFNLK.T
153	863.9200	1725.8254	1725.9512	-0.1258	1	60	0.029	1	R.GIIDAILDGSIDKAPTK.V
190	656.6100	1966.8082	1966.9783	-0.1701	0	(46)	0.52	1	K.GQVTELGAVNVMTGVYTGR.S + Oxidation (M)
191	984.4190	1966.8234	1966.9783	-0.1548	0	115	6.2e-08	1	K.GQVTELGAVNVMTGVYTGR.S + Oxidation (M)
250	989.3890	2965.1452	2965.2892	-0.1440	0	33	7.2	1	R.GIASMHCSANTNMEGTSSAIFFGLSGTGK.T + 2 Oxidation (M)

Proteins matching the same set of peptides:
gi|212692197 **Mass:** 59384 **Score:** 366 **Queries matched:** 6
hypothetical protein BACDOR_01692 [Bacteroides dorei DSM 17855]
gi|254881217 **Mass:** 59416 **Score:** 366 **Queries matched:** 6

3. gi|150009397 **Mass:** 59213 **Score:** 334 **Queries matched:** 6 **emPAI:** 0.24
phosphoenolpyruvate carboxykinase [Parabacteroides distasonis ATCC 8503]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
115	684.2770	1366.5394	1365.6677	0.8717	0	24	74	3	K.AYLVNTGWNGSGK.R
120	708.3170	1414.6194	1414.7304	-0.1109	0	93	1.2e-05	1	R.DALLENVTVAADGK.I
142	818.3880	1634.7614	1634.8515	-0.0901	0	110	2.1e-07	1	K.ELGLNSETATVFNLK.T
153	863.9200	1725.8254	1725.9149	-0.0894	0	44	0.91	2	R.GIIDAILDGSINEAPTK.K
195	993.4470	1984.8794	1985.0986	-0.2191	0	(59)	0.042	1	R.VSYPIYHIENIVKPVSK.G
196	662.6660	1984.9762	1985.0986	-0.1224	0	62	0.013	1	R.VSYPIYHIENIVKPVSK.G

4. gi|34541315 **Mass:** 59587 **Score:** 178 **Queries matched:** 4 **emPAI:** 0.24
phosphoenolpyruvate carboxykinase [Porphyromonas gingivalis W83]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
120	708.3170	1414.6194	1413.7463	0.8731	0	58	0.039	3	R.NALLENVTVDAAGK.I
153	863.9200	1725.8254	1725.9512	-0.1258	1	60	0.024	1	R.GIIDAILDGSIDKAPTK.A
195	993.4470	1984.8794	1985.0986	-0.2191	0	(59)	0.042	1	R.VSYPIYHIENIVKPVSK.A
196	662.6660	1984.9762	1985.0986	-0.1224	0	62	0.013	1	R.VSYPIYHIENIVKPVSK.A

Proteins matching the same set of peptides:
gi|188994298 **Mass:** 59614 **Score:** 178 **Queries matched:** 4
phosphoenolpyruvate carboxykinase [Porphyromonas gingivalis ATCC 33277]

5. gi|77975053 **Mass:** 59554 **Score:** 84 **Queries matched:** 3 **emPAI:** 0.11
COG1866: Phosphoenolpyruvate carboxykinase (ATP) [Yersinia frederiksenii ATCC 33641]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
151	843.3600	1684.7054	1685.7719	-1.0665	0	22	1.3e+02	2	R.LFVVDTFCGANADTR.L
195	993.4470	1984.8794	1985.0986	-0.2191	0	(59)	0.042	1	R.VSYPIYHIENIVKPVSK.A
196	662.6660	1984.9762	1985.0986	-0.1224	0	62	0.013	1	R.VSYPIYHIENIVKPVSK.A

Proteins matching the same set of peptides:
gi|77976428 **Mass:** 59498 **Score:** 84 **Queries matched:** 3
COG1866: Phosphoenolpyruvate carboxykinase (ATP) [Yersinia intermedia ATCC 29909]
gi|123444162 **Mass:** 59550 **Score:** 84 **Queries matched:** 3
phosphoenolpyruvate carboxykinase [Yersinia enterocolitica subsp. enterocolitica 8081]

6. gi|223465411 **Score:** 72 **Queries matched:** 1
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
120	708.3170	1414.6194	1413.7463	0.8731	0	72	0.0014	2	R.NALLENVTVAADGK.I

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	84	502.2610	1002.5074	1002.5862	-0.0787	0	35	9.7	1	DNIFILIR
☑	227	737.6640	2209.9702	2211.0221	-1.0519	1	33	10	1	QILCGCGSITEAQADALMKK + Oxidation (M)
☑	179	638.2790	1911.8152	1912.0054	-0.1902	0	32	13	1	AAHPGLVPPEELDALVER
☑	210	1041.9440	2081.8734	2083.0334	-1.1600	2	31	13	1	SGYERVDIVENKGEFSVR
☑	230	742.3140	2223.9202	2223.1834	0.7368	2	29	19	1	ELFGVMAAERAKGGFIVISGR + Oxidation (M)
☑	107	657.4030	1312.7914	1313.5935	-0.8021	0	29	32	1	AAGGGFAPVHGCGR
☑	194	661.3090	1980.9052	1979.9193	0.9859	1	27	41	1	MPSFAAASPQIAKSCQNR + Oxidation (M)
☑	238	1178.9000	3533.6782	3533.9259	-0.2478	2	26	56	1	ILFSRLGTPHIGPPGAFAFNVPSVRASGAITVER
☑	226	1105.9920	2209.9694	2210.0637	-0.0943	1	26	41	1	TDRGFVDSIELIAMNDATAR + Oxidation (M)
☑	58	393.3400	1176.9982	1177.5826	-0.5845	0	26	88	1	SLTEGLTSGADK
☑	130	733.8920	2198.6542	2198.0175	0.6367	2	26	90	1	VKVYMGGGGGYGDEAQRDPAR + Oxidation (M)
☑	116	692.2780	1382.5414	1382.7994	-0.2579	1	26	51	1	LSGGNQQKLIIGR
☑	219	1068.6810	3203.0212	3202.7614	0.2597	1	26	65	1	SEAGIAVLPLSARPGIYAARYIYAGIGAQVR
☑	115	684.2770	1366.5394	1365.7140	0.8254	1	26	55	1	DFIDQSSKVVTK
☑	43	366.2060	1095.5962	1094.4736	1.1226	0	26	1.6e+02	1	LADMEPMTR + 2 Oxidation (M)
☑	85	506.1470	1010.2794	1010.5760	-0.2966	0	25	46	1	DLAQPLAVGK
☑	244	905.7720	2714.2942	2715.3326	-1.0385	1	25	56	1	QTEVAIAPYDEMLFHGLRGSAPADK
☑	204	676.4300	2025.3682	2024.9916	0.3766	2	25	48	1	IGPGPGTDVEGDPIGKHHKD
☑	123	716.9830	2147.9272	2148.1440	-0.2168	2	25	1.3e+02	1	HDGHSIQKLGLPEYASKLR
☑	48	381.2650	1140.7732	1140.6404	0.1328	1	25	1.8e+02	1	WVPAKLGSQR
☑	229	1112.9580	2223.9014	2224.2388	-0.3374	2	24	58	1	LPIDVLKLDKVLVDGIEGCK
☑	105	651.8050	1301.5954	1300.7285	0.8669	2	24	98	1	KAGLAMAVADRAK
☑	96	637.3450	1272.6754	1273.7176	-1.0422	1	24	1e+02	1	LLETMGRLLGR + Oxidation (M)
☑	162	897.3670	1792.7194	1793.8465	-1.1271	0	24	73	1	MTAAVAQDTLSGADAGATK + Oxidation (M)
☑	182	965.1720	1928.3294	1928.0077	0.3217	1	24	1.3e+02	1	KQQMFVYASILESIQK + Oxidation (M)
☑	211	1042.0180	3123.0322	3123.4610	-0.4289	1	23	1.4e+02	1	WDDQAAVVASGSVMTRGMVVVPCSMGTVGR
☑	209	694.0730	2079.1972	2078.0367	1.1604	1	23	90	1	LFTRLQSAFAAMLDHADR + Oxidation (M)
☑	149	838.0810	2511.2212	2510.2660	0.9552	1	23	1.6e+02	1	MSNSPGSVPANLNGDGIRVGVVQAR + Oxidation (M)
☑	152	853.4390	2557.2952	2558.2204	-0.9252	0	23	2.1e+02	1	TGCAVSSTMGSSSSIVEATALLLSGR + Oxidation (M)

Mascot Search Results

Protein View

Match to: **gi|53715725** Score: **663**
phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6901) 10_RH23_01_1294.d\SSP (6901) 10_RH23_01_1294.mgf

Nominal mass (M_r): **59339**; Calculated pI value: **5.73**
NCBI BLAST search of gi|53715725 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60683662 from Bacteroides fragilis NCTC 9343
gi|265767330 from Bacteroides sp. 2_1_16
gi|60390334 from Bacteroides fragilis
gi|81313291 from Bacteroides fragilis NCTC 9343
gi|52218590 from Bacteroides fragilis YCH46
gi|60495096 from Bacteroides fragilis NCTC 9343
gi|263252635 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **35%**

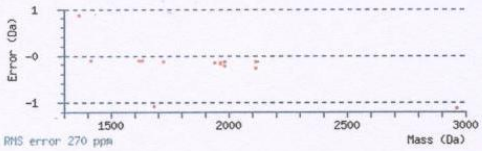
Matched peptides shown in Bold Red

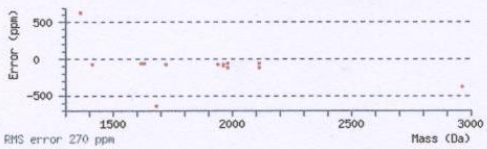
```
  1 MANLDLSKYG ITGVTEILHN PSYDVLFAEE TKPGLEGFEK GQVTELGAVN
 51 VMTGVYTGRS PKDKFFVKNE ASENSVWWTS EEYKNDNKPC SEEAWADLKA
101 KAVKELSNKR LFVVDTFCGA NEGTRMKVRF IMEVAWQAHF VTNMFIRPTA
151 EELANYGEPD FVCFNASKAK VDNYKELGLN SETATVFNLK TKEQVILNTW
201 YGGEMKKGMF SIMNYMNPLR GIASMHCSAN TDMEGTSSAI FFGLSGTGKT
251 TLSTDPKRKL IGDDEHGWDN EGVFNYEGGC YAKVINLDKE SEPDIFNAIK
301 RDALLENVTV AADGKINFAD KSVTENTRVS YPIYHIENIV KPVSKGPHAK
351 QVIFLSADAF GVLPPVSILN PEQAQYYFLS GFTAKLAGTE RGITEPTPTF
401 SACFGAAFLS LHPTKYAEEL VKKMEMTGAK AYLVNTGWNG SGKRISIKDT
451 RGIIDAILDG SIDKAPTKVI PFFDFVVPTE LPGVDPKILD PRDTYADPAQ
501 WNEKAKDLAG RFIKNFAKFT GNEAGKKLVA AGPKL
```

Show predicted peptides also

Sort Peptides By | ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
41 - 59	656.6100	1966.8082	1966.9783	-0.1701	0	K.GQVTELGAVNVMTGVYTGR.S Oxidation (M) (Ions score 46)
41 - 59	984.4190	1966.8234	1966.9783	-0.1548	0	K.GQVTELGAVNVMTGVYTGR.S Oxidation (M) (Ions score 115)
111 - 125	843.3600	1684.7054	1685.7719	-1.0665	0	R.LFVVDTFCGANEGTR.M (Ions score 32)
176 - 190	818.3880	1634.7614	1634.8515	-0.0901	0	K.ELGLNSETATVFNLK.T (Ions score 110)
208 - 220	811.3150	1620.6154	1620.7098	-0.0944	0	K.GMFSIMNYMNPLR.G 3 Oxidation (M) (Ions score 47)
221 - 249	989.3890	2965.1452	2966.2732	-1.1281	0	R.GIASMHCSANTDMEGTSSAIFFGLSGTGK.T 2 Oxidation (M) (Ions score 33)
284 - 300	972.9460	1943.8774	1944.0204	-0.1429	1	K.VINLDKESEPDIFNAIK.R (Ions score 53)
302 - 315	708.3170	1414.6194	1414.7304	-0.1109	0	R.DALLENVTVAADGK.I (Ions score 93)
329 - 345	993.4470	1984.8794	1985.0986	-0.2191	0	R.VSYPIYHIENIVKPVSK.G (Ions score 59)
329 - 345	662.6660	1984.9762	1985.0986	-0.1224	0	R.VSYPIYHIENIVKPVSK.G (Ions score 62)
431 - 443	684.2770	1366.5394	1365.6677	0.8717	0	K.AYLVNTGWNGSGK.R (Ions score 24)
452 - 468	863.9200	1725.8254	1725.9512	-0.1258	1	R.GIIDAILDGSIDKAPTK.V (Ions score 60)
469 - 487	705.9670	2114.8792	2115.1293	-0.2501	0	K.VIPFFDFVVPTELPGVDPK.I (Ions score 16)
469 - 487	1058.5150	2115.0154	2115.1293	-0.1138	0	K.VIPFFDFVVPTELPGVDPK.I (Ions score 41)






```
LOCUS       YP_101717                535 aa            linear   BCT 26-APR-2009
DEFINITION  phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46].
ACCESSION   YP_101717
VERSION     YP_101717.1  GI:53715725
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 535)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 535)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 535)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
```

```
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BAD51183.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..535
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..535
                     /product="phosphoenolpyruvate carboxykinase"
                     /EC_number="4.1.1.49"
                     /calculated_mol_wt=58897
     Region          15..532
                     /region_name="PEPCK_ATP"
                     /note="Phosphoenolpyruvate carboxykinase (PEPCK), a
                     critical gluconeogenic enzyme, catalyzes the first
                     committed step in the diversion of tricarboxylic acid
                     cycle intermediates toward gluconeogenesis. It catalyzes
                     the reversible decarboxylation and...; cd00484"
                     /db_xref="CDD:29830"
     Site            order(59,201,203,206..207,226,245..251,263..264,281,283,
                     292,328,444,447,450)
                     /site_type="active"
                     /db_xref="CDD:29830"
     Site            order(59,201,203,206..207,281,328)
                     /site_type="other"
                     /note="substrate-binding site"
                     /db_xref="CDD:29830"
     Site            order(207,226,250,263..264,281)
                     /site_type="metal-binding"
                     /db_xref="CDD:29830"
     Site            order(245..251,292,444,450)
                     /site_type="other"
                     /note="ATP binding site"
                     /db_xref="CDD:29830"
     CDS             1..535
                     /locus_tag="BF4446"
                     /coded_by="NC_006347.1:5088646..5090253"
                     /note="PEP carboxykinase; PEP carboxylase; PEPCK;
                     catalyzes the phosphorylation and decarboxylation of
                     oxaloacetate to form phosphoenolpyruvate using ATP"
                     /transl_table=11
                     /db_xref="GeneID:3085522"
```

Mascot: http://www.matrixscience.com/

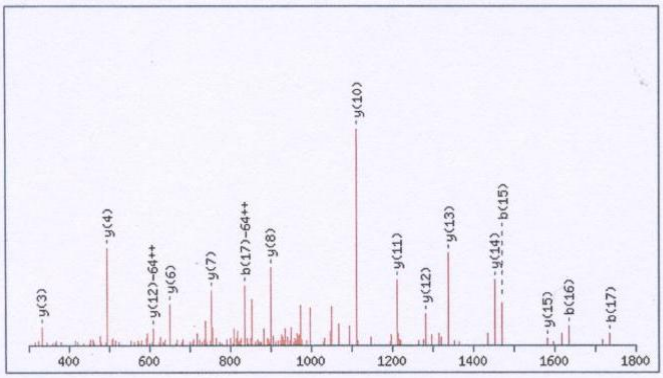
{MATRIX SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of GQVTELGAVNVMTGVYTGR
Found in gi|53715725, phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]

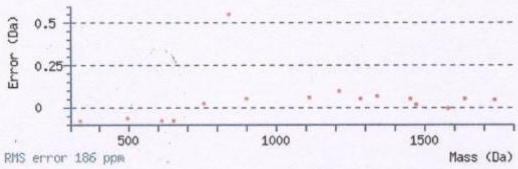
Match to Query 191: 1966.823448 from(984.419000,2+) intensity(3902720.0000)
Title: Cmpd 36, +MSn(984.70), 6.6 min
Data file D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6901) 10_RH23_01_1294.d\SSP (6901) 10_RH23_01_1294.mgf

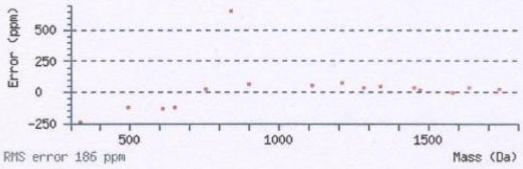
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 300 to 1800 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 1966.9783
Fixed modifications: Carboxymethyl (C)
Variable modifications:
M12 : Oxidation (M), with neutral losses 0.0000(shown in table), 63.9983
Ions Score: 115 **Expect:** 5.2e-08
Matches (Bold Red): 16/312 fragment ions using 18 most intense peaks

#	a	a^{++}	a*	a^{*++}	b	b^{++}	b*	b^{*++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	30.0338	15.5206			58.0287	29.5180			G					19
2	158.0924	79.5498	141.0659	71.0366	186.0873	93.5473	169.0608	85.0340	Q	1910.9640	955.9857	1893.9375	947.4724	18
3	257.1608	129.0840	240.1343	120.5708	285.1557	143.0815	268.1292	134.5682	V	1782.9055	891.9564	1765.8789	883.4431	17
4	358.2085	179.6079	341.1819	171.0946	386.2034	193.6053	369.1769	185.0921	T	1683.8371	842.4222	1666.8105	833.9089	16
5	487.2511	244.1292	470.2245	235.6159	515.2460	258.1266	498.2195	249.6134	E	**1582.7894**	791.8983	1565.7628	783.3851	15
6	600.3352	300.6712	583.3086	292.1579	628.3301	314.6687	611.3035	306.1554	L	**1453.7468**	727.3770	1436.7202	718.8638	14
7	657.3566	329.1819	640.3301	320.6687	685.3515	343.1794	668.3250	334.6661	G	**1340.6627**	670.8350	1323.6362	662.3217	13
8	728.3937	364.7005	711.3672	356.1872	756.3886	378.6980	739.3621	370.1847	A	**1283.6413**	642.3243	1266.6147	633.8110	12
9	827.4621	414.2347	810.4356	405.7214	855.4571	428.2322	838.4305	419.7189	V	**1212.6041**	606.8057	1195.5776	598.2924	11
10	941.5051	471.2562	924.4785	462.7429	969.5000	485.2536	952.4734	476.7404	N	**1113.5357**	557.2715	1096.5092	548.7582	10
11	1040.5735	520.7904	1023.5469	512.2771	1068.5684	534.7878	1051.5419	526.2746	V	999.4928	500.2500	982.4662	491.7368	9
12	1187.6089	594.3081	1170.5823	585.7948	1215.6038	608.3055	1198.5773	599.7923	M	**900.4244**	450.7158	883.3978	442.2026	8
13	1288.6566	644.8319	1271.6300	636.3186	1316.6515	658.8294	1299.6249	650.3161	T	**753.3890**	377.1981	736.3624	368.6849	7
14	1345.6780	673.3427	1328.6515	664.8294	1373.6729	687.3401	1356.6464	678.8268	G	**652.3413**	326.6743	635.3148	318.1610	6
15	1444.7464	722.8769	1427.7199	714.3636	**1472.7414**	736.8743	1455.7148	728.3610	V	595.3198	298.1636	578.2933	289.6503	5
16	1607.8098	804.4085	1590.7832	795.8953	**1635.8047**	818.4060	1618.7781	809.8927	Y	**496.2514**	248.6293	479.2249	240.1161	4
17	1708.8575	854.9324	1691.8309	846.4191	**1736.8524**	868.9298	1719.8258	860.4165	T	**333.1881**	167.0977	316.1615	158.5844	3
18	1765.8789	883.4431	1748.8524	874.9298	1793.8738	897.4406	1776.8473	888.9273	G	232.1404	116.5738	215.1139	108.0606	2
19									R	175.1190	88.0631	158.0924	79.5498	1





NCBI **BLAST** search of GQVTELGAVNVMTGVYTGR
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
115.5	1966.9783	-0.1548	GQVTELGAVNVMTGVYTGR
27.0	1966.0054	0.8180	SAEMAARAAEHVAGLPTLR
24.8	1965.9843	0.8391	AFNRSQPNMDVAGLFKR
18.9	1966.0861	0.7374	VVDAERALARFSGVALHR
18.5	1967.1350	-0.3115	AAEKERLMVLVAGLNVVR
17.8	1966.1211	0.7024	SLSPELKVTPEGRATLLR
15.6	1967.9160	-1.0925	MGQYNFDQIVDRTHTK
14.5	1965.9948	0.8286	VYTGGKADWYAAGLPLER
14.4	1967.0952	-0.2718	ARFVVELVQAVAEAVGPGR
13.6	1967.0298	-0.2064	MNKHPEFLLNKIENPK

Mascot: http://www.matrixscience.com/

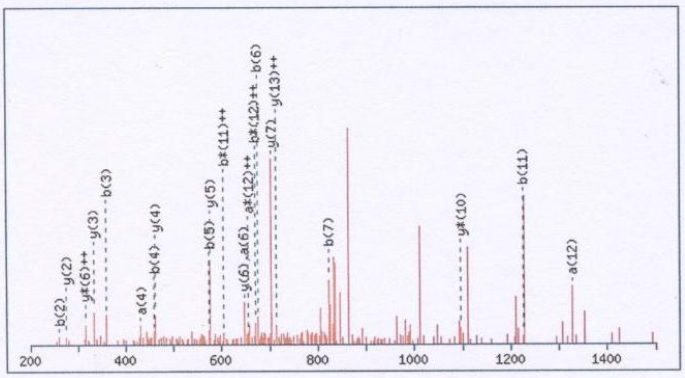
Mascot Search Results

Peptide View

MS/MS Fragmentation of LFVVDTFCGANEGTR
Found in gi|53715725, phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]

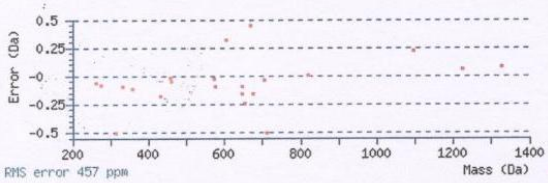
Match to Query 151: 1684.705448 from(843.360000,2+) intensity(5734853.0000)
Title: Cmpd 46, +MSn(843.76), 6.9 min
Data file D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6901) 10_RH23_01_1294.d\SSP (6901) 10_RH23_01_1294.mgf

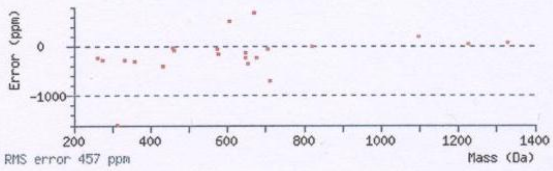
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 200 to 1500 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 1685.7719
Fixed modifications: Carboxymethyl (C)
Ions Score: 32 **Expect:** 13
Matches (Bold Red): 22/128 fragment ions using 61 most intense peaks

#	a	a^{++}	a*	a^{*++}	b	b^{++}	b*	b^{*++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	86.0964	43.5519			114.0913	57.5493			L					15
2	233.1648	117.0861			**261.1598**	131.0835			F	1573.6951	787.3512	1556.6686	778.8379	14
3	332.2333	166.6203			**360.2282**	180.6177			V	1426.6267	**713.8170**	1409.6002	705.3037	13
4	**431.3017**	216.1545			**459.2966**	230.1519			V	1327.5583	664.2828	1310.5318	655.7695	12
5	546.3286	273.6679			**574.3235**	287.6654			D	1228.4899	614.7486	1211.4633	606.2353	11
6	**647.3763**	324.1918			**675.3712**	338.1892			T	1113.4630	557.2351	**1096.4364**	548.7218	10
7	794.4447	397.7260			**822.4396**	411.7234			F	1012.4153	506.7113	995.3887	498.1980	9
8	955.4594	478.2333			983.4543	492.2308			C	865.3469	433.1771	848.3203	424.6638	8
9	1012.4808	506.7441			1040.4757	520.7415			G	**704.3322**	352.6697	687.3056	344.1565	7
10	1083.5179	542.2626			1111.5129	556.2601			A	**647.3107**	324.1590	630.2842	**315.6457**	6
11	1197.5609	599.2841	1180.5343	590.7708	**1225.5558**	613.2815	1208.5292	**604.7683**	N	**576.2736**	288.6404	559.2471	280.1272	5
12	**1326.6035**	663.8054	1309.5769	**655.2921**	1354.5984	677.8028	1337.5718	**669.2896**	E	**462.2307**	231.6190	445.2041	223.1057	4
13	1383.6249	692.3161	1366.5984	683.8028	1411.6198	706.3136	1394.5933	697.8003	G	**333.1881**	167.0977	316.1615	158.5844	3
14	1484.6726	742.8399	1467.6461	734.3267	1512.6675	756.8374	1495.6410	748.3241	T	**276.1666**	138.5870	259.1401	130.0737	2
15									R	175.1190	88.0631	158.0924	79.5498	1





NCBI **BLAST** search of LFVVDTFCGANEGTR
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
32.2	1685.7719	-1.0665	LFVVDTFCGANEGTR
22.2	1685.7719	-1.0665	LFVVDTFCGANADTR
21.9	1683.9083	0.7971	FEGVSKIYSTDVVLK
21.4	1685.8083	-1.1029	NFVFDERMSEVVSK
19.4	1685.8413	-1.1359	VIFDQFNFTDTIAR
18.6	1685.7865	-1.0810	ACQIMAGHLKENADK
17.4	1684.8056	-0.1002	GNSLSDTYIEVFANR
16.9	1684.7978	-0.0924	MIVEGTGAYTTVNGEK
16.4	1683.8502	0.8553	AENMKTTVNSFSVLK
16.3	1685.8698	-1.1644	KALFDYEMLKDVSK

Mascot: http://www.matrixscience.com/

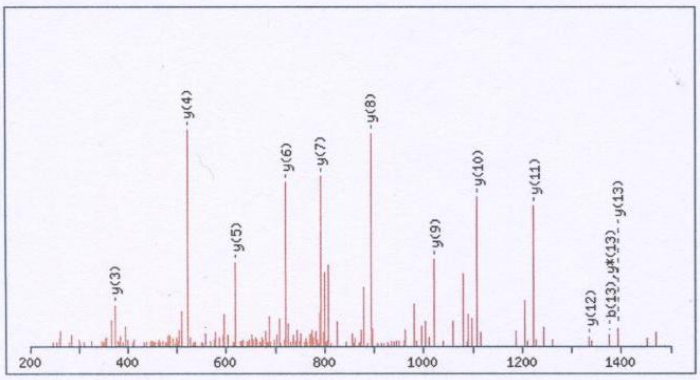
{MATRIX SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of ELGLNSETATVFNLK
Found in gi|53715725, phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]

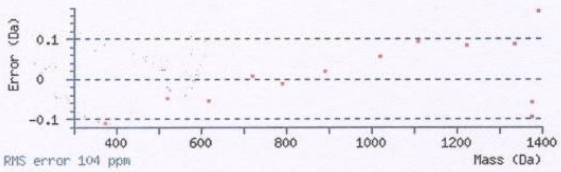
Match to Query 142: 1634.761448 from(818.388000,2+) intensity(10020720.0000)
Title: Cmpd 60, +MSn(818.83), 7.6 min
Data file D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6901) 10_RH23_01_1294.d\SSP (6901) 10_RH23_01_1294.mgf

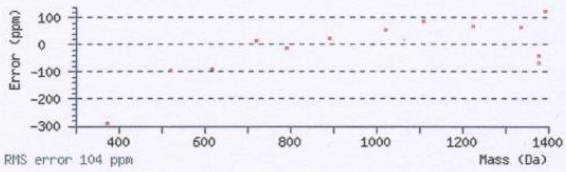
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 200 to 1500 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 1634.8515
Fixed modifications: Carboxymethyl (C)
Ions Score: 110 **Expect:** 2.1e-07
Matches (Bold Red): 13/152 fragment ions using 16 most intense peaks

#	a	a^{++}	a*	a^{*++}	b	b^{++}	b*	b^{*++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	102.0550	51.5311			130.0499	65.5286			E					15
2	215.1390	108.0731			243.1339	122.0706			L	1506.8162	753.9118	1489.7897	745.3985	14
3	272.1605	136.5839			300.1554	150.5813			G	**1393.7322**	697.3697	**1376.7056**	688.8564	13
4	385.2445	193.1259			413.2395	207.1234			L	**1336.7107**	668.8590	1319.6842	660.3457	12
5	499.2875	250.1474	482.2609	241.6341	527.2824	264.1448	510.2558	255.6316	N	**1223.6266**	612.3170	1206.6001	603.8037	11
6	586.3195	293.6634	569.2930	285.1501	614.3144	307.6608	597.2879	299.1476	S	**1109.5837**	555.2955	1092.5572	546.7822	10
7	715.3621	358.1847	698.3355	349.6714	743.3570	372.1821	726.3305	363.6689	E	**1022.5517**	511.7795	1005.5251	503.2662	9
8	816.4098	408.7085	799.3832	400.1953	844.4047	422.7060	827.3781	414.1927	T	**893.5091**	447.2582	876.4825	438.7449	8
9	887.4469	444.2271	870.4203	435.7138	915.4418	458.2245	898.4153	449.7113	A	**792.4614**	396.7343	775.4349	388.2211	7
10	988.4946	494.7509	971.4680	486.2376	1016.4895	508.7484	999.4629	500.2351	T	**721.4243**	361.2158	704.3978	352.7025	6
11	1087.5630	544.2851	1070.5364	535.7719	1115.5579	558.2826	1098.5313	549.7693	V	**620.3766**	310.6919	603.3501	302.1787	5
12	1234.6314	617.8193	1217.6048	609.3061	1262.6263	631.8168	1245.5998	623.3035	F	**521.3082**	261.1577	504.2817	252.6445	4
13	1348.6743	674.8408	1331.6478	666.3275	**1376.6692**	688.8383	1359.6427	680.3250	N	**374.2398**	187.6235	357.2132	179.1103	3
14	1461.7584	731.3828	1444.7318	722.8696	1489.7533	745.3803	1472.7268	736.8670	L	260.1969	130.6021	243.1703	122.0888	2
15									K	147.1128	74.0600	130.0863	65.5468	1





NCBI **BLAST** search of ELGLNSETATVFNLK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
110.3	1634.8515	-0.0901	ELGLNSETATVFNLK
22.3	1635.8688	-1.1073	IGLSEMLTALSMKAR
20.5	1635.8356	-1.0741	EYVEVTAETVEVIR
18.9	1633.8522	0.9092	TELEIRTATLSASDK
18.6	1633.6937	1.0678	AAVGARDTCNPCNAR
18.0	1634.8562	-0.0948	FNDDKTAMVRALVR
18.0	1633.8610	0.9005	NLNIALFGATGMIGSR
18.0	1634.9178	-0.1563	TLKIALFGATGMIGSR
17.8	1635.7425	-0.9811	EMVFITWDSHMPK
16.9	1634.8562	-0.0948	KNGLGGFSVRQLCAK

Mascot: http://www.matrixscience.com/

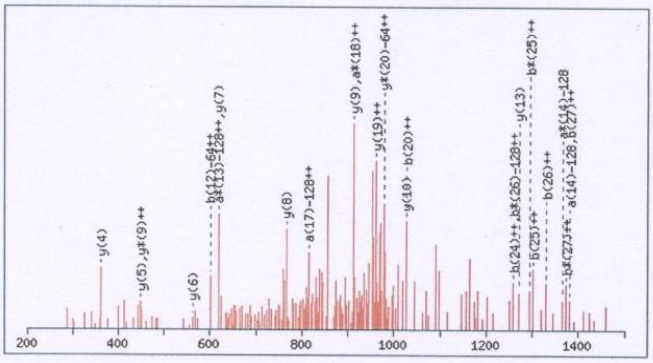
MATRIX SCIENCE Mascot Search Results

Peptide View

MS/MS Fragmentation of GIASMHCSANTDMEGTSSAIFFGLSGTGK
Found in gi|53715725, phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]

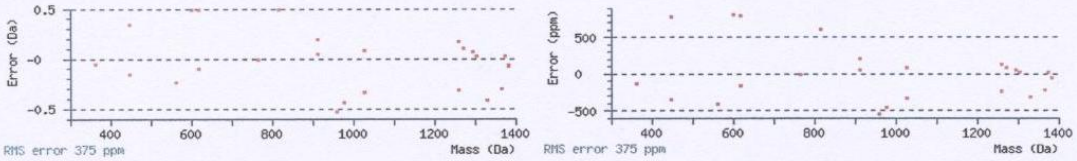
Match to Query 250: 2965.145172 from(989.389000,3+) intensity(3604278.0000)
Title: Cmpd 55, +MSn(989.77), 7.4 min
Data file D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6901) 10_RH23_01_1294.d\SSP (6901) 10_RH23_01_1294.mgf

Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 200 to 1500 Da Full range




```
Monoisotopic mass of neutral peptide Mr(calc): 2966.2732
Fixed modifications: Carboxymethyl (C)
Variable modifications:
M5     : Oxidation (M), with neutral losses 0.0000(shown in table), 63.9983
M13    : Oxidation (M), with neutral losses 0.0000(shown in table), 63.9983
Ions Score: 33  Expect: 6
Matches (Bold Red): 25/520 fragment ions using 30 most intense peaks
```

#	a	a++	a*	a*++	b	b++	b*	b*++	Seq.	y	y++	y*	y*++	#
1	30.0338	15.5206			58.0287	29.5180			G					29
2	143.1179	72.0626			171.1128	86.0600			I	2910.2590	1455.6332	2893.2325	1447.1199	28
3	214.1550	107.5811			242.1499	121.5786			A	2797.1750	1399.0911	2780.1484	1390.5779	27
4	301.1870	151.0972			329.1819	165.0946			S	2726.1379	1363.5726	2709.1113	1355.0593	26
5	448.2224	224.6149			476.2173	238.6123			M	2639.1058	1320.0566	2622.0793	1311.5433	25
6	585.2813	293.1443			613.2763	307.1418			H	2492.0704	1246.5389	2475.0439	1238.0256	24
7	746.2960	373.6516			774.2909	387.6491			C	2355.0115	1178.0094	2337.9850	1169.4961	23
8	833.3280	417.1677			861.3230	431.1651			S	2193.9969	1097.5021	2176.9703	1088.9888	22
9	904.3652	452.6862			932.3601	466.6837			A	2106.9648	1053.9861	2089.9383	1045.4728	21
10	1018.4081	509.7077	1001.3815	501.1944	1046.4030	523.7051	1029.3764	515.1919	N	2035.9277	1018.4675	2018.9012	1009.9542	20
11	1119.4558	560.2315	1102.4292	551.7182	1147.4507	574.2290	1130.4241	565.7157	T	1921.8848	**961.4460**	1904.8582	952.9328	19
12	1234.4827	617.7450	1217.4562	609.2317	1262.4776	631.7424	1245.4511	623.2292	D	1820.8371	910.9222	1803.8106	902.4089	18
13	1381.5181	691.2627	1364.4916	682.7494	1409.5130	705.2601	1392.4865	696.7469	M	1705.8102	853.4087	1688.7836	844.8954	17
14	1510.5607	755.7840	1493.5342	747.2707	1538.5556	769.7814	1521.5291	761.2682	E	1558.7748	779.8910	1541.7482	771.3777	16
15	1567.5822	784.2947	1550.5556	775.7814	1595.5771	798.2922	1578.5505	789.7789	G	1429.7322	715.3697	1412.7056	706.8564	15
16	1668.6298	834.8186	1651.6033	826.3053	1696.6248	848.8160	1679.5982	840.3027	T	1372.7107	686.8590	1355.6842	678.3457	14
17	1755.6619	878.3346	1738.6353	869.8213	1783.6568	892.3320	1766.6302	883.8188	S	**1271.6630**	636.3352	1254.6365	627.8219	13
18	1842.6939	921.8506	1825.6673	**913.3373**	1870.6888	935.8480	1853.6623	927.3348	S	1184.6310	592.8191	1167.6045	584.3059	12
19	1913.7310	957.3691	1896.7045	948.8559	1941.7259	971.3666	1924.6994	962.8533	A	1097.5990	549.3031	1080.5724	540.7898	11
20	2026.8151	1013.9112	2009.7885	1005.3979	2054.8100	**1027.9086**	2037.7834	1019.3954	I	**1026.5619**	513.7846	1009.5353	505.2713	10
21	2173.8835	1087.4454	2156.8569	1078.9321	2201.8784	1101.4428	2184.8519	1092.9296	F	**913.4778**	457.2425	896.4512	**448.7293**	9
22	2320.9519	1160.9796	2303.9254	1152.4663	2348.9468	1174.9770	2331.9203	1166.4638	F	**766.4094**	383.7083	749.3828	375.1951	8
23	2377.9734	1189.4903	2360.9468	1180.9770	2405.9683	1203.4878	2388.9417	1194.9745	G	**619.3410**	310.1741	602.3144	301.6608	7
24	2491.0574	1246.0324	2474.0309	1237.5191	2519.0523	**1260.0298**	2502.0258	1251.5165	L	**562.3195**	281.6634	545.2930	273.1501	6
25	2578.0895	1289.5484	2561.0629	1281.0351	2606.0844	**1303.5458**	2589.0578	**1295.0326**	S	**449.2354**	225.1214	432.2089	216.6081	5
26	2635.1109	1318.0591	2618.0844	1309.5458	2663.1058	**1332.0566**	2646.0793	1323.5433	G	**362.2034**	181.6053	345.1769	173.0921	4
27	2736.1586	1368.5829	2719.1321	1360.0697	2764.1535	**1382.5804**	2747.1270	**1374.0671**	T	305.1819	153.0946	288.1554	144.5813	3
28	2793.1801	1397.0937	2776.1535	1388.5804	2821.1750	1411.0911	2804.1484	1402.5779	G	204.1343	102.5708	187.1077	94.0575	2
29									K	147.1128	74.0600	130.0863	65.5468	1



NCBI **BLAST** search of GIASMHCSANTDMEGTSSAIFFGLSGTGK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
32.7	2966.2732	-1.1281	GIASMHCSANTDMEGTSSAIFFGLSGTGK
32.7	2965.2892	-0.1440	GIASMHCSANTNMEGTSSAIFFGLSGTGK
32.0	2965.2416	-0.0965	DVMSMHCSANVGEDGDTTLFFGLSGTGK
23.5	2965.3586	-0.2135	ADVLPMHCAANVDDAGNTALFFGLSGTGK
21.3	2964.4916	0.6535	RGVFPMHASANVGKEGDVALFFGLSGTGK
20.8	2964.5266	0.6186	LNLTAPVTPEQGKINYISIAMAGYSADK
18.6	2966.3273	-1.1821	LVEVMENYIASLDEEALAGMAGSSNNSH
17.3	2964.3900	0.7551	NEDFFLFNLIFFCKFVKMSDMQK
17.1	2964.4949	0.6502	SEHQMAMLDQYAGHLNLLKTHVSLTK
16.9	2964.6331	0.5121	IAIHHAEGMLQIGEIPVIITASSKHRK

Mascot: http://www.matrixscience.com/

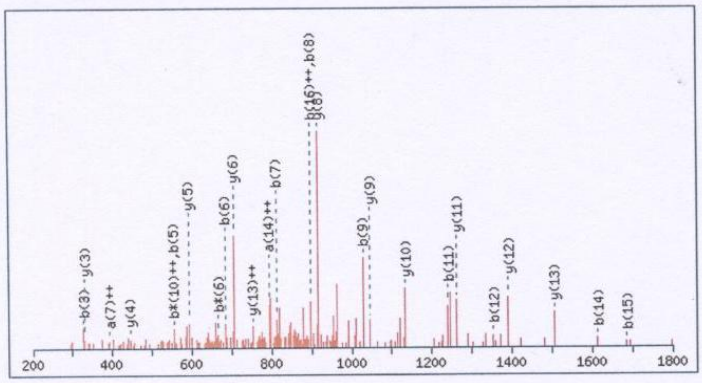
{MATRIX SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of VINLDKESEPDIFNAIK
Found in gi|53715725, phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]

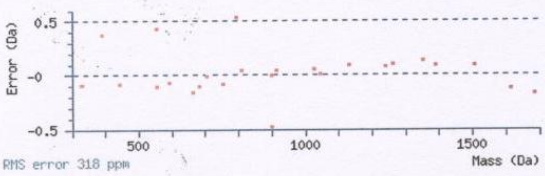
Match to Query 186: 1943.877448 from(972.946000,2+) intensity(3662039.0000)
Title: Cmpd 62, +MSn(973.22), 7.7 min
Data file D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6901) 10_RH23_01_1294.d\SSP (6901) 10_RH23_01_1294.mgf

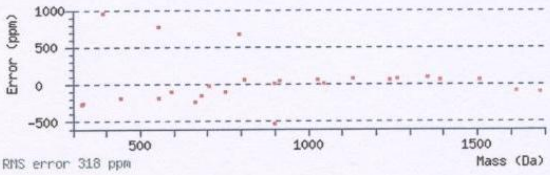
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 200 to 1800 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 1944.0204
Fixed modifications: Carboxymethyl (C)
Ions Score: 53 **Expect:** 0.12
Matches (Bold Red): 26/184 fragment ions using 52 most intense peaks

#	a	a^{++}	a*	a^{*++}	b	b^{++}	b*	b^{*++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	72.0808	36.5440			100.0757	50.5415			V					17
2	185.1648	93.0861			213.1598	107.0835			I	1845.9593	923.4833	1828.9327	914.9700	16
3	299.2078	150.1075	282.1812	141.5942	**327.2027**	164.1050	310.1761	155.5917	N	1732.8752	866.9412	1715.8487	858.4280	15
4	412.2918	206.6496	395.2653	198.1363	440.2867	220.6470	423.2602	212.1337	L	1618.8323	809.9198	1601.8057	801.4065	14
5	527.3188	264.1630	510.2922	255.6498	**555.3137**	278.1605	538.2871	269.6472	D	**1505.7482**	**753.3777**	1488.7217	744.8645	13
6	655.4137	328.2105	638.3872	319.6972	**683.4087**	342.2080	**666.3821**	333.6947	K	**1390.7213**	695.8643	1373.6947	687.3510	12
7	784.4563	**392.7318**	767.4298	384.2185	**812.4512**	406.7293	795.4247	398.2160	E	**1262.6263**	631.8168	1245.5998	623.3035	11
8	871.4884	436.2478	854.4618	427.7345	**899.4833**	450.2453	882.4567	441.7320	S	**1133.5837**	567.2955	1116.5572	558.7822	10
9	1000.5309	500.7691	983.5044	492.2558	**1028.5259**	514.7666	1011.4993	506.2533	E	**1046.5517**	523.7795	1029.5251	515.2662	9
10	1097.5837	549.2955	1080.5572	540.7822	1125.5786	563.2930	1108.5521	**554.7797**	P	**917.5091**	459.2582	900.4825	450.7449	8
11	1212.6107	606.8090	1195.5841	598.2957	**1240.6056**	620.8064	1223.5790	612.2931	D	820.4563	410.7318	803.4298	402.2185	7
12	1325.6947	663.3510	1308.6682	654.8377	**1353.6896**	677.3485	1336.6631	668.8352	I	**705.4294**	353.2183	688.4028	344.7051	6
13	1472.7631	736.8852	1455.7366	728.3719	1500.7581	750.8827	1483.7315	742.3694	F	**592.3453**	296.6763	575.3188	288.1630	5
14	1586.8061	**793.9067**	1569.7795	785.3934	**1614.8010**	807.9041	1597.7744	799.3909	N	**445.2769**	223.1421	428.2504	214.6288	4
15	1657.8432	829.4252	1640.8166	820.9120	**1685.8381**	843.4227	1668.8115	834.9094	A	**331.2340**	166.1206	314.2074	157.6074	3
16	1770.9272	885.9673	1753.9007	877.4540	1798.9222	**899.9647**	1781.8956	891.4514	I	260.1969	130.6021	243.1703	122.0888	2
17									K	147.1128	74.0600	130.0863	65.5468	1





NCBI **BLAST** search of VINLDKESEPDIFNAIK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
52.5	1944.0204	-0.1429	VINLDKESEPDIFNAIK
26.7	1944.0204	-0.1429	VINLDKDSEPEIFNAIK
16.6	1944.0713	-0.1939	IRVADKETALEALQMLK
10.9	1942.9095	0.9680	FGGIDKEGIYIDENAMR
9.7	1945.0417	-1.1642	ILMSMPITPGPALLAFNK
9.7	1943.9839	-0.1065	KYAQAEEPEESPLAAALK
8.6	1943.8935	-0.0160	NATSPFEATGSLYPMAGSK
8.2	1943.9921	-0.1146	RQAAQVMESGVVAGGMIPK
7.9	1943.8275	0.0500	SDTVTMPTDEMREAMAK
7.8	1943.9921	-0.1146	KIPICGQDDVIIRNMR

Mascot: http://www.matrixscience.com/

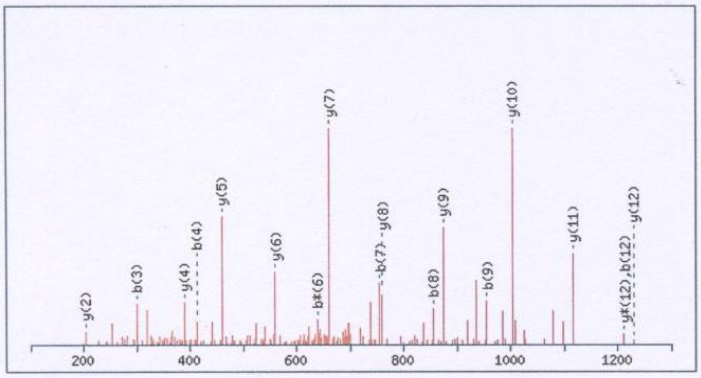
Mascot Search Results

Peptide View

MS/MS Fragmentation of DALLENVTVAADGK
Found in gi|53715725, phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]

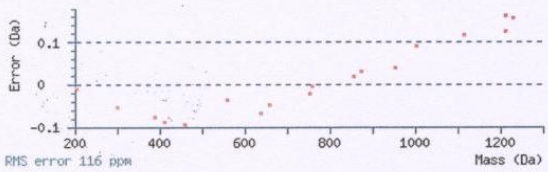
Match to Query 120: 1414.619448 from(708.317000,2+) intensity(4539975.0000)
Title: Cmpd 11, +MSn(708.71), 5.6 min
Data file D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6901) 10_RH23_01_1294.d\SSP (6901) 10_RH23_01_1294.mgf

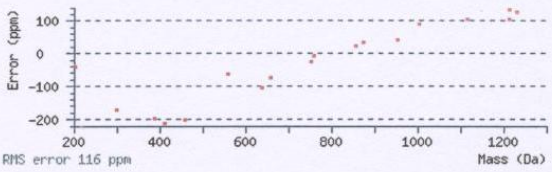
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 100 to 1300 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 1414.7304
Fixed modifications: Carboxymethyl (C)
Ions Score: 93 **Expect:** 1.2e-05
Matches (Bold Red): 18/136 fragment ions using 21 most intense peaks

#	a	a^{++}	a*	a^{*++}	b	b^{++}	b*	b^{*++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	88.0393	44.5233			116.0342	58.5207			D					14
2	159.0764	80.0418			187.0713	94.0393			A	1300.7107	650.8590	1283.6842	642.3457	13
3	272.1605	136.5839			**300.1554**	150.5813			L	**1229.6736**	615.3404	**1212.6470**	606.8272	12
4	385.2445	193.1259			**413.2395**	207.1234			L	**1116.5895**	558.7984	1099.5630	550.2851	11
5	514.2871	257.6472			542.2821	271.6447			E	**1003.5055**	502.2564	986.4789	493.7431	10
6	628.3301	314.6687	611.3035	306.1554	656.3250	328.6661	**639.2984**	320.1529	N	**874.4629**	437.7351	857.4363	429.2218	9
7	727.3985	364.2029	710.3719	355.6896	**755.3934**	378.2003	738.3668	369.6871	V	**760.4199**	380.7136	743.3934	372.2003	8
8	828.4462	414.7267	811.4196	406.2134	**856.4411**	428.7242	839.4145	420.2109	T	**661.3515**	331.1794	644.3250	322.6661	7
9	927.5146	464.2609	910.4880	455.7476	**955.5095**	478.2584	938.4829	469.7451	V	**560.3039**	280.6556	543.2773	272.1423	6
10	998.5517	499.7795	981.5251	491.2662	1026.5466	513.7769	1009.5201	505.2637	A	**461.2354**	231.1214	444.2089	222.6081	5
11	1069.5888	535.2980	1052.5623	526.7848	1097.5837	549.2955	1080.5572	540.7822	A	**390.1983**	195.6028	373.1718	187.0895	4
12	1184.6157	592.8115	1167.5892	584.2982	**1212.6107**	606.8090	1195.5841	598.2957	D	319.1612	160.0842	302.1347	151.5710	3
13	1241.6372	621.3222	1224.6107	612.8090	1269.6321	635.3197	1252.6056	626.8064	G	**204.1343**	102.5708	187.1077	94.0575	2
14									K	147.1128	74.0600	130.0863	65.5468	1





NCBI **BLAST** search of DALLENVTVAADGK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
93.1	1414.7304	-0.1109	DALLENVTVAADGK
72.4	1413.7463	0.8731	NALLENVTVAADGK
58.0	1413.7463	0.8731	NALLENVTVDAAGK
30.1	1414.7303	-0.1109	GIELENIDAADKK
25.9	1414.7164	-0.0970	GNQIEGRISEGQK
24.8	1414.7668	-0.1473	VTVIDGVASLAENK
24.0	1414.6875	-0.0680	DNGMGISVQELPR
22.0	1414.7667	-0.1473	EGLLVENLSVSQK
22.0	1414.6762	-0.0568	GELMQEAVPAGEGK
18.9	1414.8395	-0.2201	VSIIVGVSKDLTGK

Mascot: http://www.matrixscience.com/

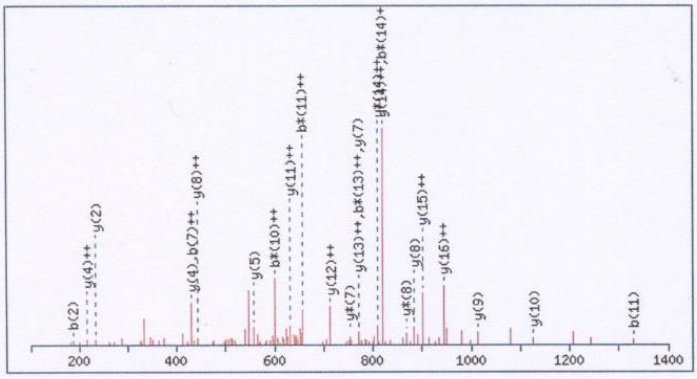
{MATRIX SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of VSYPIYHIENIVKPVSK
Found in gi|53715725, phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]

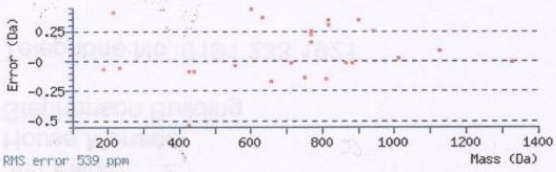
Match to Query 196: 1984.976172 from(662.666000,3+) intensity(4638027.0000)
Title: Cmpd 53, +MSn(663.74), 7.3 min
Data file D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6901) 10_RH23_01_1294.d\SSP (6901) 10_RH23_01_1294.mgf

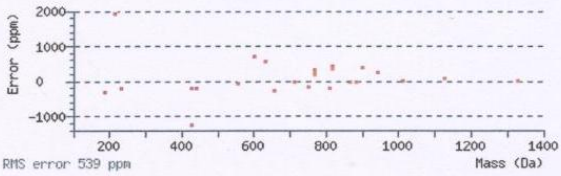
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 100 to 1400 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 1985.0986
Fixed modifications: Carboxymethyl (C)
Ions Score: 62 **Expect:** 0.013
Matches (Bold Red): 25/156 fragment ions using 35 most intense peaks

#	a	a^{++}	a*	a^{*++}	b	b^{++}	b*	b^{*++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	72.0808	36.5440			100.0757	50.5415			V					17
2	159.1128	80.0600			**187.1077**	94.0575			S	1887.0375	**944.0224**	1870.0109	935.5091	16
3	322.1761	161.5917			350.1710	175.5892			Y	1800.0054	**900.5064**	1782.9789	891.9931	15
4	419.2289	210.1181			447.2238	224.1155			P	1636.9421	**818.9747**	1619.9156	**810.4614**	14
5	532.3130	266.6601			560.3079	280.6576			I	1539.8893	**770.4483**	1522.8628	761.9350	13
6	695.3763	348.1918			723.3712	362.1892			Y	1426.8053	**713.9063**	1409.7787	705.3930	12
7	832.4352	416.7212			860.4301	**430.7187**			H	1263.7419	**632.3746**	1246.7154	623.8613	11
8	945.5193	473.2633			973.5142	487.2607			I	**1126.6830**	563.8452	1109.6565	555.3319	10
9	1074.5619	537.7846			1102.5568	551.7820			E	**1013.5990**	507.3031	996.5724	498.7898	9
10	1188.6048	594.8060	1171.5782	586.2928	1216.5997	608.8035	1199.5732	**600.2902**	N	**884.5564**	**442.7818**	**867.5298**	434.2686	8
11	1301.6888	651.3481	1284.6623	642.8348	**1329.6838**	665.3455	1312.6572	**656.8322**	I	**770.5135**	385.7604	**753.4869**	377.2471	7
12	1400.7573	700.8823	1383.7307	692.3690	1428.7522	714.8797	1411.7256	706.3665	V	657.4294	329.2183	640.4028	320.7051	6
13	1528.8522	764.9298	1511.8257	756.4165	1556.8471	778.9272	1539.8206	**770.4139**	K	**558.3610**	279.6841	541.3344	271.1709	5
14	1625.9050	813.4561	1608.8784	804.9429	1653.8999	827.4536	1636.8734	**818.9403**	P	**430.2660**	**215.6366**	413.2395	207.1234	4
15	1724.9734	862.9903	1707.9469	854.4771	1752.9683	876.9878	1735.9418	868.4745	V	333.2132	167.1103	316.1867	158.5970	3
16	1812.0054	906.5064	1794.9789	897.9931	1840.0003	920.5038	1822.9738	911.9905	S	**234.1448**	117.5761	217.1183	109.0628	2
17									K	147.1128	74.0600	130.0863	65.5468	1





NCBI **BLAST** search of VSYPIYHIENIVKPVSK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
62.3	1985.0986	-0.1224	VSYPIYHIENIVKPVSK
32.6	1985.0986	-0.1224	VSYPIYHIDNIVKPISK
24.3	1985.1018	-0.1256	VDQSRASRTGGTGLGLAIVK
24.1	1984.0993	0.8769	HDIPPLLADISAALKPKAD
22.8	1984.1146	0.8616	VSYPIYHIQNIVKPVSK
22.8	1984.0782	0.8980	VSYPIYHIQNIVQPVSK
22.6	1986.0157	-1.0395	VTLEDDNILINGEKVDSL
21.2	1986.0291	-1.0530	YIGNIPVSMHAVMGAVRR
20.7	1985.2481	-0.2720	WFIGVAIIILIIFIINK
20.1	1985.0080	-0.0319	SMNLGPLFEPFELHNLK

Mascot: http://www.matrixscience.com/

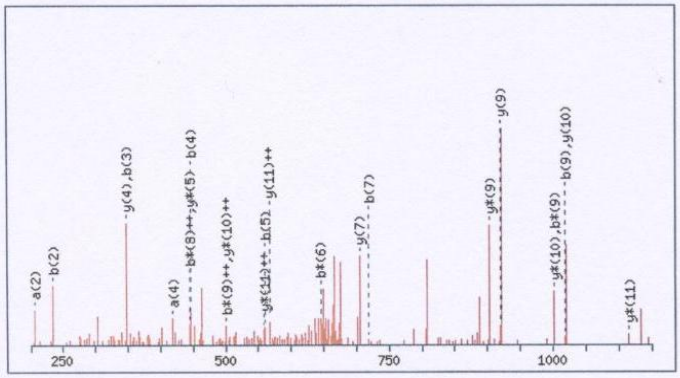
{MATRIX SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of AYLVNTGWNGSGK
Found in gi|53715725, phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]

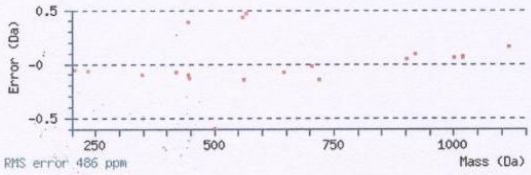
Match to Query 115: 1366.539448 from(684.277000,2+) intensity(1224761.0000)
Title: Cmpd 4, +MSn(684.95), 5.3 min
Data file D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6901) 10_RH23_01_1294.d\SSP (6901) 10_RH23_01_1294.mgf

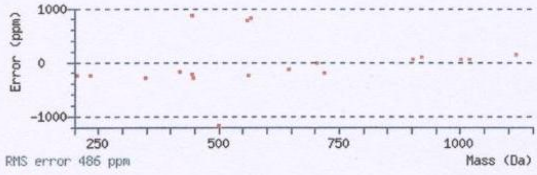
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 200 to 1150 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 1365.6677
Fixed modifications: Carboxymethyl (C)
Ions Score: 24 **Expect:** 74
Matches (Bold Red): 23/128 fragment ions using 42 most intense peaks

#	a	a++	a*	a*++	b	b++	b*	b*++	Seq.	y	y++	y*	y*++	#
1	44.0495	22.5284			72.0444	36.5258			A					13
2	**207.1128**	104.0600			**235.1077**	118.0575			Y	1295.6379	648.3226	1278.6113	639.8093	12
3	320.1969	160.6021			**348.1918**	174.5995			L	1132.5745	**566.7909**	**1115.5480**	**558.2776**	11
4	**419.2653**	210.1363			**447.2602**	224.1337			V	**1019.4905**	510.2489	**1002.4639**	**501.7356**	10
5	533.3082	267.1577	516.2817	258.6445	**561.3031**	281.1552	544.2766	272.6419	N	**920.4221**	460.7147	**903.3955**	452.2014	9
6	634.3559	317.6816	617.3293	309.1683	662.3508	331.6790	**645.3243**	323.1658	T	806.3791	403.6932	789.3526	395.1799	8
7	691.3774	346.1923	674.3508	337.6790	**719.3723**	360.1898	702.3457	351.6765	G	**705.3315**	353.1694	688.3049	344.6561	7
8	877.4567	439.2320	860.4301	430.7187	905.4516	453.2294	888.4250	**444.7162**	W	648.3100	324.6586	631.2835	316.1454	6
9	991.4996	496.2534	974.4730	487.7402	**1019.4945**	510.2509	**1002.4680**	**501.7376**	N	462.2307	231.6190	**445.2041**	223.1057	5
10	1048.5211	524.7642	1031.4945	516.2509	1076.5160	538.7616	1059.4894	530.2483	G	**348.1878**	174.5975	331.1612	166.0842	4
11	1135.5531	568.2802	1118.5265	559.7669	1163.5480	582.2776	1146.5215	573.7644	S	291.1663	146.0868	274.1397	137.5735	3
12	1192.5745	596.7909	1175.5480	588.2776	1220.5695	610.7884	1203.5429	602.2751	G	204.1343	102.5708	187.1077	94.0575	2
13									K	147.1128	74.0600	130.0863	65.5468	1





NCBI **BLAST** search of AYLVNTGWNGSGK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence

25.7	1365.7140	0.8254	DFIDQSSKVVTK
25.3	1365.7504	0.7891	YVTVSSPSKALSK
24.4	1365.6677	0.8717	AYLVNTGWNGSGK
23.8	1365.6525	0.8870	FSIVDRGDENSK
19.5	1366.7391	-0.1996	MSIDSRRFVIK
18.9	1366.7642	-0.2248	YALLKLVTGSMR
18.9	1366.6299	-0.0905	FSPDSCDRVKR
18.7	1365.7405	0.7990	HFGKPGAEPLVSK
18.7	1366.7496	-0.2102	YALALFQIATEK
18.7	1366.7496	-0.2102	YALALFQLATEK

Mascot: http://www.matrixscience.com/

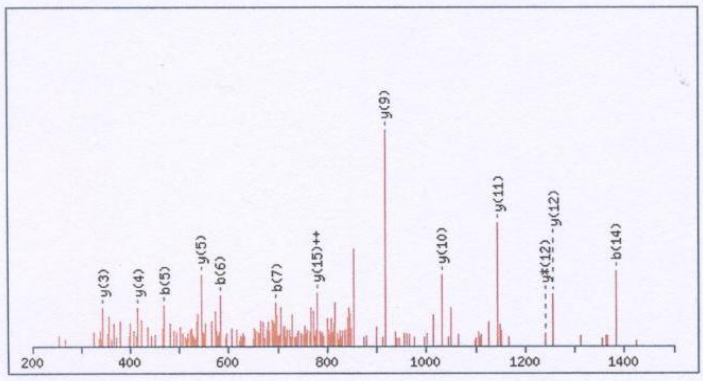
Mascot Search Results

Peptide View

MS/MS Fragmentation of GIIDAILDGSIDKAPTK
Found in gi|53715725, phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]

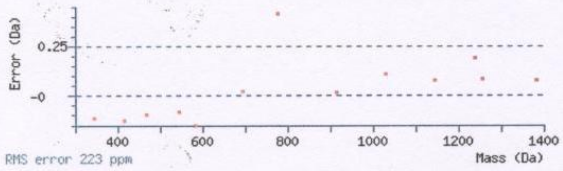
Match to Query 153: 1725.825448 from(863.920000,2+) intensity(2365091.0000)
Title: Cmpd 72, +MSn(864.71), 8.1 min
Data file D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6901) 10_RH23_01_1294.d\SSP (6901) 10_RH23_01_1294.mgf

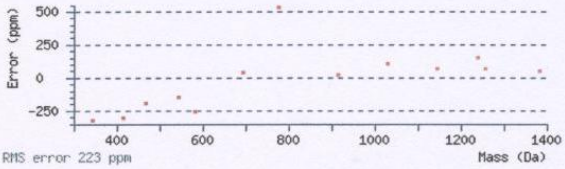
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 200 to 1500 Da Full range



Monoisotopic mass of neutral peptide Mr(calc): 1725.9512
Fixed modifications: Carboxymethyl (C)
Ions Score: 60 **Expect:** 0.024
Matches (Bold Red): 13/144 fragment ions using 16 most intense peaks

#	a	a^{++}	a*	a^{*++}	b	b^{++}	b*	b^{*++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	30.0338	15.5206			58.0287	29.5180			G					17
2	143.1179	72.0626			171.1128	86.0600			I	1669.9371	835.4722	1652.9105	826.9589	16
3	256.2020	128.6046			284.1969	142.6021			I	1556.8530	**778.9301**	1539.8265	770.4169	15
4	371.2289	186.1181			399.2238	200.1155			D	1443.7690	722.3881	1426.7424	713.8748	14
5	442.2660	221.6366			**470.2609**	235.6341			A	1328.7420	664.8746	1311.7155	656.3614	13
6	555.3501	278.1787			**583.3450**	292.1761			I	**1257.7049**	629.3561	**1240.6783**	620.8428	12
7	668.4341	334.7207			**696.4291**	348.7182			L	**1144.6208**	572.8141	1127.5943	564.3008	11
8	783.4611	392.2342			811.4560	406.2316			D	**1031.5368**	516.2720	1014.5102	507.7587	10
9	840.4825	420.7449			868.4775	434.7424			G	**916.5098**	458.7585	899.4833	450.2453	9
10	927.5146	464.2609			955.5095	478.2584			S	859.4884	430.2478	842.4618	421.7345	8
11	1040.5986	520.8030			1068.5936	534.8004			I	772.4563	386.7318	755.4298	378.2185	7
12	1155.6256	578.3164			1183.6205	592.3139			D	659.3723	330.1898	642.3457	321.6765	6
13	1283.7205	642.3639	1266.6940	633.8506	1311.7155	656.3614	1294.6889	647.8481	K	**544.3453**	272.6763	527.3188	264.1630	5
14	1354.7577	677.8825	1337.7311	669.3692	**1382.7526**	691.8799	1365.7260	683.3666	A	**416.2504**	208.6288	399.2238	200.1155	4
15	1451.8104	726.4088	1434.7839	717.8956	1479.8053	740.4063	1462.7788	731.8930	P	**345.2132**	173.1103	328.1867	164.5970	3
16	1552.8581	776.9327	1535.8316	768.4194	1580.8530	790.9301	1563.8265	782.4169	T	248.1605	124.5839	231.1339	116.0706	2
17									K	147.1128	74.0600	130.0863	65.5468	1





NCBI **BLAST** search of GIIDAILDGSIDKAPTK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc):	Delta	Sequence
59.6	1725.9512	-0.1258	GIIDAILDGSIDKAPTK
43.9	1725.9149	-0.0894	GIIDAILDGSINEAPTK
21.9	1726.8923	-1.0669	NNIEIIDATCPVVLR
21.8	1725.8291	-0.0036	DGPGAGKLMADLMTHGR
21.8	1726.9326	-1.1071	DLNQLLDAVSSRLQR
18.7	1726.7655	-0.9400	MSDPTVMIHDPEPAR
17.7	1725.9341	-0.1087	ILNEIYWVSTYVVK
17.7	1725.9824	-0.1570	MPIQLLRHAPGAPGLR
17.7	1725.9447	-0.1193	IVEGALLAAGKPMDVAR
17.2	1725.8645	-0.0391	AVDNSSDAIASILAHSR

Mascot: http://www.matrixscience.com/

Mascot Search Results

Peptide View

MS/MS Fragmentation of **VIPFFDFVVPTELPGVDPK**
Found in **gi|53715725**, phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]

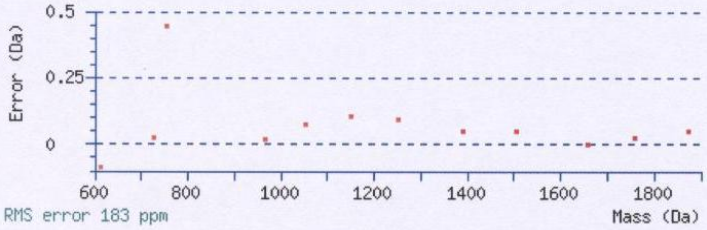
Match to Query 216: 2115.015448 from(1058.515000,2+) intensity(2057610.0000)
Title: Cmpd 180, +MSn(1059.16), 11.6 min
Data file D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (6901) 10_RH23_01_1294.d\SSP (6901) 10_RH23_01_1294.mgf

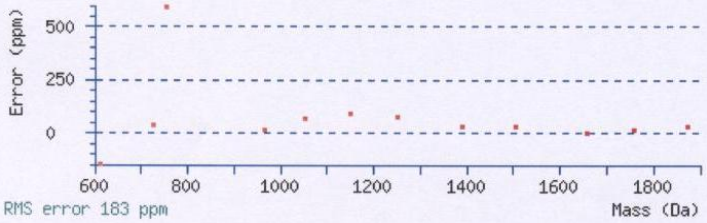
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 300 to 1900 Da Full range

Monoisotopic mass of neutral peptide Mr(calc): 2115.1293
Fixed modifications: Carboxymethyl (C)
Ions Score: 41 **Expect:** 1.9
Matches (Bold Red): 12/144 fragment ions using 18 most intense peaks

#	a	a^{++}	b	b^{++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	72.0808	36.5440	100.0757	50.5415	V					19
2	185.1648	93.0861	213.1598	107.0835	I	2017.0681	1009.0377	2000.0415	1000.5244	18
3	282.2176	141.6124	310.2125	155.6099	P	1903.9840	952.4957	1886.9575	943.9824	17
4	429.2860	215.1466	457.2809	229.1441	F	1806.9313	903.9693	1789.9047	895.4560	16
5	576.3544	288.6809	604.3493	302.6783	F	1659.8629	830.4351	1642.8363	821.9218	15
6	691.3814	346.1943	719.3763	360.1918	D	1512.7944	756.9009	1495.7679	748.3876	14
7	838.4498	419.7285	866.4447	433.7260	F	1397.7675	699.3874	1380.7409	690.8741	13
8	937.5182	469.2627	**965.5131**	483.2602	V	**1250.6991**	625.8532	1233.6725	617.3399	12
9	1036.5866	518.7969	1064.5815	532.7944	V	**1151.6307**	576.3190	1134.6041	567.8057	11
10	1133.6394	567.3233	1161.6343	581.3208	P	**1052.5623**	526.7848	1035.5357	518.2715	10
11	1234.6871	617.8472	1262.6820	631.8446	T	955.5095	478.2584	938.4829	469.7451	9
12	1363.7297	682.3685	**1391.7246**	696.3659	E	854.4618	427.7345	837.4353	419.2213	8

13	1476.8137	738.9105	1504.8086	752.9080	L	725.4192	363.2132	708.3927	354.7000	7
14	1573.8665	787.4369	1601.8614	801.4343	P	612.3352	306.6712	595.3086	298.1579	6
15	1630.8879	815.9476	1658.8829	829.9451	G	515.2824	258.1448	498.2558	249.6316	5
16	1729.9564	865.4818	1757.9513	879.4793	V	458.2609	229.6341	441.2344	221.1208	4
17	1844.9833	922.9953	1872.9782	936.9927	D	359.1925	180.0999	342.1660	171.5866	3
18	1942.0361	971.5217	1970.0310	985.5191	P	244.1656	122.5864	227.1390	114.0731	2
19					K	147.1128	74.0600	130.0863	65.5468	1





NCBI **BLAST** search of VIPFFDFVVPTELPGVDPK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

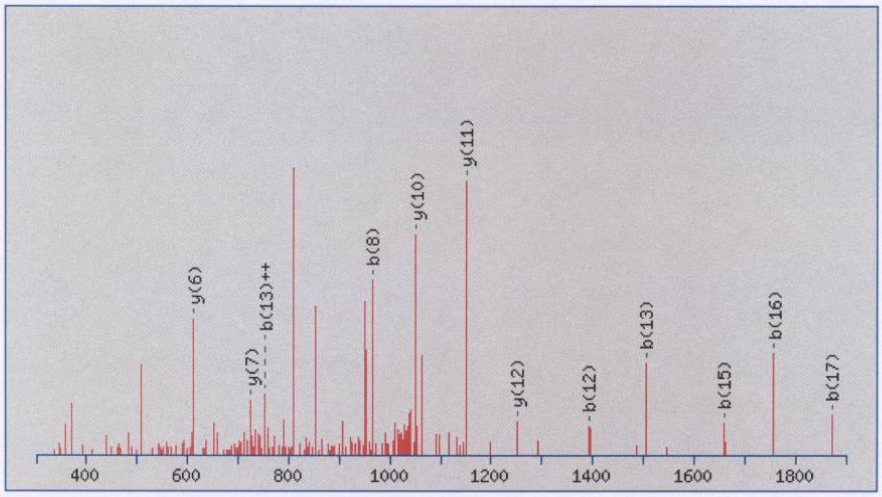
All matches to this query

Score	Mr(calc):	Delta	Sequence
41.0	2115.1293	-0.1138	VIPFFDFVVPTELPGVDPK
18.9	2115.2013	-0.1859	IKEKMPDIPVIVISAFTSK
18.0	2115.2892	-0.2737	IPISAEVAAIVRQRGLVVPK
15.8	2114.9513	0.0641	DENGVPMRDHLQGPEYMK
15.4	2115.1940	-0.1785	ILDAAQIVDIVSDFVTLRK
15.4	2115.1940	-0.1785	ILDAAQIVEVVSDFVTLRK
15.0	2115.2489	-0.2335	LLPKMGTIPEVVILDPPRK
14.7	2115.1589	-0.1435	LVDVGKIPHPGRGANFVDPK
14.1	2114.9757	0.0398	VSGNETFDDVSLSVQFESR
13.9	2114.9903	0.0252	KDEGHDVVVTVSAMAGETDR

Mascot: http://www.matrixscience.com/

Bacteroides fragilis SSP 6901- Peptide sequences

VIPFFDFVVPTELPGVDPK- Ion score 41

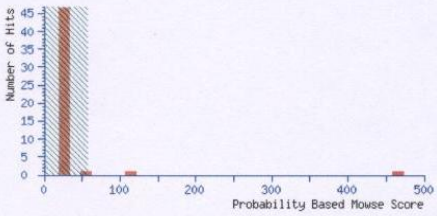


Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 7301
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (7301) 10_RJ12_01_1384.d\SSP (7301) 10_RJ12_01_1384.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 16:22:43 GMT
Protein hits    : gi|143945   Fe-superoxide dismutase [Bacteroides fragilis]
                  gi|29346065 superoxide dismutase [Fe] [Bacteroides thetaiotaomicron VPI-5482]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐

Select All | Select None | Search Selected | ☐ **Error tolerant**

1. gi|143945 **Mass:** 21736 **Score:** 466 **Queries matched:** 37 **emPAI:** 1.36
   Fe-superoxide dismutase [Bacteroides fragilis]
   ☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	40	545.6590	1089.3034	1089.4767	-0.1733	0	35	6.8	1	R.DFGSFENFK.K
☑	78	772.7580	1543.5014	1543.7922	-0.2908	0	(69)	0.0026	1	K.LWEIIDWDVVEK.R
☑	79	515.5170	1543.5292	1543.7922	-0.2631	0	(35)	6.2	1	K.LWEIIDWDVVEK.R
☑	80	772.7800	1543.5454	1543.7922	-0.2468	0	84	7.8e-05	1	K.LWEIIDWDVVEK.R
☑	81	772.7810	1543.5474	1543.7922	-0.2448	0	(70)	0.0019	1	K.LWEIIDWDVVEK.R
☑	82	772.7880	1543.5614	1543.7922	-0.2308	0	(54)	0.096	1	K.LWEIIDWDVVEK.R
☑	83	772.7930	1543.5714	1543.7922	-0.2208	0	(62)	0.025	1	K.LWEIIDWDVVEK.R
☑	84	772.8010	1543.5874	1543.7922	-0.2048	0	(66)	0.0057	1	K.LWEIIDWDVVEK.R
☑	85	772.8190	1543.6234	1543.7922	-0.1688	0	(54)	0.1	1	K.LWEIIDWDVVEK.R
☑	86	772.8390	1543.6634	1543.7922	-0.1288	0	(35)	8.3	1	K.LWEIIDWDVVEK.R
☑	181	1120.9470	2239.8794	2240.0902	-0.2108	0	68	0.0027	1	K.EFNAASVGLFGSGWAWLSVDK.D
☑	193	792.5530	2374.6372	2375.1757	-0.5385	0	(34)	5.9	1	K.HLQTYVNNLNSLVPGTEYEGK.T
☑	195	792.5820	2374.7242	2375.1757	-0.4515	0	(40)	1.4	1	K.HLQTYVNNLNSLVPGTEYEGK.T
☑	196	792.5970	2374.7692	2375.1757	-0.4065	0	(27)	29	1	K.HLQTYVNNLNSLVPGTEYEGK.T
☑	197	792.6000	2374.7782	2375.1757	-0.3975	0	(37)	3.1	1	K.HLQTYVNNLNSLVPGTEYEGK.T
☑	198	792.6190	2374.8352	2375.1757	-0.3405	0	50	0.17	1	K.HLQTYVNNLNSLVPGTEYEGK.T
☑	200	792.6670	2374.9792	2375.1757	-0.1965	0	(27)	37	1	K.HLQTYVNNLNSLVPGTEYEGK.T
☑	212	847.6150	2539.8232	2540.2336	-0.4104	1	(48)	0.22	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L
☑	213	847.6320	2539.8742	2540.2336	-0.3594	1	(37)	3.2	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L
☑	214	847.6330	2539.8772	2540.2336	-0.3564	1	(49)	0.2	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L
☑	215	847.6490	2539.9252	2540.2336	-0.3084	1	(63)	0.0078	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L
☑	216	847.6530	2539.9372	2540.2336	-0.2964	1	(53)	0.082	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L
☑	217	847.7230	2540.1472	2540.2336	-0.0864	1	(66)	0.0056	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L
☑	109	847.9690	2540.8852	2540.2336	0.6516	1	83	0.00021	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L
☑	223	878.6390	2632.8952	2633.3125	-0.4174	0	(72)	0.00094	1	K.LPYANNALEPVISQQTIDYHYGK.H
☑	224	878.6430	2632.9072	2633.3125	-0.4054	0	72	0.00092	1	K.LPYANNALEPVISQQTIDYHYGK.H
☑	225	878.6580	2632.9522	2633.3125	-0.3604	0	(54)	0.064	1	K.LPYANNALEPVISQQTIDYHYGK.H
☑	226	878.6660	2632.9762	2633.3125	-0.3364	0	(61)	0.011	1	K.LPYANNALEPVISQQTIDYHYGK.H
☑	227	878.6780	2633.0122	2633.3125	-0.3004	0	(55)	0.051	1	K.LPYANNALEPVISQQTIDYHYGK.H
☑	228	878.9610	2633.8612	2633.3125	0.5486	0	(53)	0.069	1	K.LPYANNALEPVISQQTIDYHYGK.H
☑	229	878.9610	2633.8612	2633.3125	0.5486	0	(43)	0.74	1	K.LPYANNALEPVISQQTIDYHYGK.H
☑	230	878.9770	2633.9092	2633.3125	0.5966	0	(57)	0.028	1	K.LPYANNALEPVISQQTIDYHYGK.H
☑	231	878.9950	2633.9632	2633.3125	0.6506	0	(62)	0.0095	1	K.LPYANNALEPVISQQTIDYHYGK.H

☑	240	923.3140	2766.9202	2767.3758	-0.4556	0	(49)	0.17	1	R.AGLKPLLGFDVWEHAYYLDYQNR.R
☑	241	923.3210	2766.9412	2767.3758	-0.4346	0	73	0.00066	1	R.AGLKPLLGFDVWEHAYYLDYQNR.R
☑	242	923.3360	2766.9862	2767.3758	-0.3896	0	(30)	13	1	R.AGLKPLLGFDVWEHAYYLDYQNR.R
☑	243	923.3410	2767.0012	2767.3758	-0.3746	0	(50)	0.15	1	R.AGLKPLLGFDVWEHAYYLDYQNR.R

Proteins matching the same set of peptides:
gi|53713818 **Mass:** 23124 **Score:** 466 **Queries matched:** 37
superoxide dismutase [Bacteroides fragilis YCH46]
gi|60682036 **Mass:** 21758 **Score:** 466 **Queries matched:** 37
superoxide dismutase [Fe] [Bacteroides fragilis NCTC 9343]

---

2. gi|29346065 **Mass:** 21762 **Score:** 123 **Queries matched:** 8 **emPAI:** 0.33
superoxide dismutase [Fe] [Bacteroides thetaiotaomicron VPI-5482]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
	40	545.6590	1089.3034	1089.4767	-0.1733	0	35	6.8	1	R.DFGSFENFK.K
	193	792.5530	2374.6372	2375.1757	-0.5385	0	(34)	5.9	1	K.HLQTYVNNLNSLVPGTEYEGK.T
	195	792.5820	2374.7242	2375.1757	-0.4515	0	(40)	1.4	1	K.HLQTYVNNLNSLVPGTEYEGK.T
	196	792.5970	2374.7692	2375.1757	-0.4065	0	(27)	29	1	K.HLQTYVNNLNSLVPGTEYEGK.T
	197	792.6000	2374.7782	2375.1757	-0.3975	0	(37)	3.1	1	K.HLQTYVNNLNSLVPGTEYEGK.T
	198	792.6190	2374.8352	2375.1757	-0.3405	0	50	0.17	1	K.HLQTYVNNLNSLVPGTEYEGK.T
	200	792.6670	2374.9792	2375.1757	-0.1965	0	(27)	37	1	K.HLQTYVNNLNSLVPGTEYEGK.T
☑	245	928.6340	2782.8802	2783.3707	-0.4906	0	38	2.2	1	R.AGLKPLLGFDVWEHSYYLDYQNR.R

Proteins matching the same set of peptides:
gi|153806587 **Mass:** 23126 **Score:** 123 **Queries matched:** 8
hypothetical protein BACCAC_00857 [Bacteroides caccae ATCC 43185]
gi|237719542 **Mass:** 22532 **Score:** 123 **Queries matched:** 8
superoxide dismutase [Bacteroides sp. 2_2_4]
gi|253568517 **Mass:** 23094 **Score:** 123 **Queries matched:** 8
conserved hypothetical protein [Bacteroides sp. 1_1_6]
gi|160884385 **Mass:** 23105 **Score:** 121 **Queries matched:** 8
hypothetical protein BACOVA_02363 [Bacteroides ovatus ATCC 8483]
gi|255692463 **Mass:** 23142 **Score:** 121 **Queries matched:** 8
superoxide dismutase [Bacteroides finegoldii DSM 17565]
gi|260174595 **Mass:** 23119 **Score:** 121 **Queries matched:** 8
superoxide dismutase (Fe) [Bacteroides sp. D2]

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	89	788.7800	1575.5454	1575.8449	-0.2995	1	39	2.8	1	WLEAKDFLLWQK
☑	61	693.0460	1384.0774	1383.6228	0.4547	0	38	3.4	1	SMLPYDEIASDK + Oxidation (M)
☑	105	828.7830	1655.5514	1655.9498	-0.3984	0	32	12	1	LFTVLPELTPAQLSK
☑	156	692.5960	2074.7662	2074.2150	0.5512	1	31	15	1	IANEPPILQIISEVKAALR
☑	194	792.5760	2374.7062	2375.2341	-0.5279	0	28	24	1	GSNAPLLDFAIAVGAVAAMIVMGR + 2 Oxidation (M)
☑	44	563.7460	1125.4774	1125.5198	-0.0424	0	28	41	1	EMFGAIAMPK + 2 Oxidation (M)
☑	150	1004.8860	2007.7574	2008.0068	-0.2493	1	27	39	1	WAVFGSGWGGAFGVEVARR
☑	239	922.6520	2764.9342	2765.4745	-0.5404	1	26	35	1	VLINQVPGGMISNLANQLKEQGALDK + Oxidation (M)
☑	60	693.0310	1384.0474	1384.5816	-0.5342	0	26	55	1	CEETGIDYEIR
☑	176	737.9140	2210.7202	2209.9698	0.7503	1	25	48	1	IADDYMPRGQWSGSSSDNPK
☑	37	507.1970	1012.3794	1012.4899	-0.1105	1	25	79	1	YMIKADEK + Oxidation (M)
☑	22	908.8810	907.8737	907.5603	0.3134	1	25	54	1	LRLLAVSH
☑	165	696.8470	2087.5192	2087.2177	0.3015	2	25	51	1	MKVGIVGGGPAGLYSAILLKK + Oxidation (M)
☑	119	884.2970	1766.5794	1766.8257	-0.2463	1	25	1e+02	1	MIEEGEAGFLRADSSR
☑	87	779.7580	2336.2522	2335.0824	1.1697	2	25	1.5e+02	1	MDQTSSDVQKIDLMKYFDR + Oxidation (M)
☑	153	1011.2920	3030.8542	3029.6814	1.1728	2	24	1.1e+02	1	YTEQFLTAAAGTSFAAPLLIHKAARLLR
☑	187	1142.7100	3425.1082	3425.6715	-0.5633	1	24	1.1e+02	1	GYGPLLVFGDSDGDAWMLRDFADTAVGVIVNR
☑	52	631.2090	1890.6052	1891.0965	-0.4913	1	24	1.5e+02	1	KTFITGAIILMAAGLVTR + Oxidation (M)
☑	161	692.9330	2075.7772	2074.9676	0.8095	2	24	72	1	NYLCFSINGANRTMKNR + Oxidation (M)
☑	219	858.9390	2573.7952	2574.3402	-0.5450	0	24	59	1	VGLAHGSVSGILQEGIDSSNPIAPTR
☑	235	903.6510	2707.9312	2708.4385	-0.5073	1	24	58	1	VEVLRDGASTIYGSDAIAGVINFITK
☑	116	585.8940	1754.6602	1754.8873	-0.2271	1	24	89	1	EGMTTGDISKALFLTR + Oxidation (M)
☑	218	858.2970	2571.8692	2571.2421	0.6271	0	24	65	1	MQIAAPADPMVNGAQSDSASVAATIR
☑	55	673.7860	1345.5574	1346.6652	-1.1078	1	23	1.3e+02	1	KELAPMFPENR + Oxidation (M)
☑	70	734.3110	1466.6074	1466.8093	-0.2018	1	23	1.2e+02	1	NEPLKITTELPGR
☑	59	691.0200	2070.0382	2071.0256	-0.9874	0	23	2.1e+02	1	MTDSTQTPDIPLLDAVQAR
☑	88	780.7600	1559.5054	1558.8501	0.6554	1	23	97	1	LANEILREMVLSR + Oxidation (M)
☑	159	692.6780	2075.0122	2074.0405	0.9717	1	23	1.2e+02	1	IYPVSGMSKIGSGSYTLNGK + Oxidation (M)
☑	158	692.6560	2074.9462	2074.0591	0.8871	2	23	1.2e+02	1	LLTIMDELRVGCPWDKK
☑	123	905.3180	2712.9322	2713.5420	-0.6099	2	23	1.6e+02	1	MKKALAVVLAVVQLAMCVVALSGVAR + Oxidation (M)
☑	47	594.8990	1187.7834	1187.6033	0.1801	0	23	2.6e+02	1	ELQEQAASLLS
☑	126	605.9000	1814.6782	1814.9738	-0.2956	2	23	1e+02	1	VADLGKVVDTLADEKSR
☑	24	910.9870	909.9797	910.4695	-0.4897	1	22	91	1	MYKGVAAR + Oxidation (M)
☑	160	692.7580	2075.2522	2076.1149	-0.8628	2	22	1.2e+02	1	REILIALGKMAAVDYQNR + Oxidation (M)
☑	96	800.5800	2398.7182	2398.3584	0.3598	0	22	2e+02	1	AATAGAVAGGVSGGLAGGVVAVGILYIK
☑	173	1083.2770	3246.8092	3246.4194	0.3898	1	22	1.8e+02	1	GGSLFCASGDDQLQAGDSVYIVCARDNVDR

Mascot Search Results

Protein View

Match to: **gi|60682036** Score: **466**
superoxide dismutase [Fe] [Bacteroides fragilis NCTC 9343]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP (7301) 10_RJ12_01_1384.d\SSP (7301) 10_RJ12_01_1384.mgf

Nominal mass (M_r): **21758**; Calculated pI value: **6.07**
NCBI BLAST search of gi|60682036 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis NCTC 9343
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|253565789 from Bacteroides sp. 3_2_5
gi|265764163 from Bacteroides sp. 2_1_16
gi|55977805 from Bacteroides fragilis
gi|60493470 from Bacteroides fragilis NCTC 9343
gi|251946068 from Bacteroides sp. 3_2_5
gi|263256771 from Bacteroides sp. 2_1_16

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **58%**

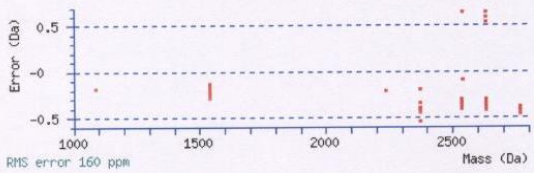
Matched peptides shown in **Bold Red**

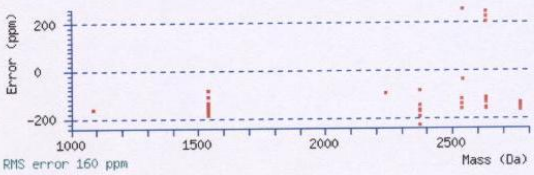
```
  1 MTYEMPKLPY ANNALEPVIS QQTIDYHYGK HLQTYVNNLN SLVPGTEYEG
 51 KTVEAIVASA PDGAIFNNAG QVLNHTLYFL QFAPKPAKNE PAGKLGEAIK
101 RDFGSFENFK KEFNAASVGL FGSGWAWLSV DKDGKLHITK EPNGSNPVRA
151 GLKPLLGFDV WEHAYYLDYQ NRRADHVNKL WEIIDWDVVE KRL
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence	
8 - 30	878.6390	2632.8952	2633.3125	-0.4174	0	K.LPYANNALEPVISQQTIDYHYGK.H	(Ions score 72)
8 - 30	878.6430	2632.9072	2633.3125	-0.4054	0	K.LPYANNALEPVISQQTIDYHYGK.H	(Ions score 72)
8 - 30	878.6580	2632.9522	2633.3125	-0.3604	0	K.LPYANNALEPVISQQTIDYHYGK.H	(Ions score 54)
8 - 30	878.6660	2632.9762	2633.3125	-0.3364	0	K.LPYANNALEPVISQQTIDYHYGK.H	(Ions score 61)
8 - 30	878.6780	2633.0122	2633.3125	-0.3004	0	K.LPYANNALEPVISQQTIDYHYGK.H	(Ions score 55)
8 - 30	878.9610	2633.8612	2633.3125	0.5486	0	K.LPYANNALEPVISQQTIDYHYGK.H	(Ions score 53)
8 - 30	878.9610	2633.8612	2633.3125	0.5486	0	K.LPYANNALEPVISQQTIDYHYGK.H	(Ions score 43)
8 - 30	878.9770	2633.9092	2633.3125	0.5966	0	K.LPYANNALEPVISQQTIDYHYGK.H	(Ions score 57)
8 - 30	878.9950	2633.9632	2633.3125	0.6506	0	K.LPYANNALEPVISQQTIDYHYGK.H	(Ions score 62)
31 - 51	792.5530	2374.6372	2375.1757	-0.5385	0	K.HLQTYVNNLNSLVPGTEYEGK.T	(Ions score 34)
31 - 51	792.5820	2374.7242	2375.1757	-0.4515	0	K.HLQTYVNNLNSLVPGTEYEGK.T	(Ions score 40)
31 - 51	792.5970	2374.7692	2375.1757	-0.4065	0	K.HLQTYVNNLNSLVPGTEYEGK.T	(Ions score 27)
31 - 51	792.6000	2374.7782	2375.1757	-0.3975	0	K.HLQTYVNNLNSLVPGTEYEGK.T	(Ions score 37)
31 - 51	792.6190	2374.8352	2375.1757	-0.3405	0	K.HLQTYVNNLNSLVPGTEYEGK.T	(Ions score 50)
31 - 51	792.6670	2374.9792	2375.1757	-0.1965	0	K.HLQTYVNNLNSLVPGTEYEGK.T	(Ions score 27)
102 - 110	545.6590	1089.3034	1089.4767	-0.1733	0	R.DFGSFENFK.K	(Ions score 35)
112 - 132	1120.9470	2239.8794	2240.0902	-0.2108	0	K.EFNAASVGLFGSGWAWLSVDK.D	(Ions score 68)
112 - 135	847.6150	2539.8232	2540.2336	-0.4104	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L	(Ions score 48)
112 - 135	847.6320	2539.8742	2540.2336	-0.3594	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L	(Ions score 37)
112 - 135	847.6330	2539.8772	2540.2336	-0.3564	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L	(Ions score 49)
112 - 135	847.6490	2539.9252	2540.2336	-0.3084	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L	(Ions score 63)
112 - 135	847.6530	2539.9372	2540.2336	-0.2964	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L	(Ions score 53)
112 - 135	847.7230	2540.1472	2540.2336	-0.0864	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L	(Ions score 66)
112 - 135	847.9690	2540.8852	2540.2336	0.6516	1	K.EFNAASVGLFGSGWAWLSVDKDGK.L	(Ions score 83)
150 - 172	923.3140	2766.9202	2767.3758	-0.4556	0	R.AGLKPLLGFDVWEHAYYLDYQNR.R	(Ions score 49)
150 - 172	923.3210	2766.9412	2767.3758	-0.4346	0	R.AGLKPLLGFDVWEHAYYLDYQNR.R	(Ions score 73)
150 - 172	923.3360	2766.9862	2767.3758	-0.3896	0	R.AGLKPLLGFDVWEHAYYLDYQNR.R	(Ions score 30)
150 - 172	923.3410	2767.0012	2767.3758	-0.3746	0	R.AGLKPLLGFDVWEHAYYLDYQNR.R	(Ions score 50)
180 - 191	772.7580	1543.5014	1543.7922	-0.2908	0	K.LWEIIDWDVVEK.R	(Ions score 69)
180 - 191	515.5170	1543.5292	1543.7922	-0.2631	0	K.LWEIIDWDVVEK.R	(Ions score 35)
180 - 191	772.7800	1543.5454	1543.7922	-0.2468	0	K.LWEIIDWDVVEK.R	(Ions score 84)
180 - 191	772.7810	1543.5474	1543.7922	-0.2448	0	K.LWEIIDWDVVEK.R	(Ions score 70)
180 - 191	772.7880	1543.5614	1543.7922	-0.2308	0	K.LWEIIDWDVVEK.R	(Ions score 54)
180 - 191	772.7930	1543.5714	1543.7922	-0.2208	0	K.LWEIIDWDVVEK.R	(Ions score 62)
180 - 191	772.8010	1543.5874	1543.7922	-0.2048	0	K.LWEIIDWDVVEK.R	(Ions score 66)
180 - 191	772.8190	1543.6234	1543.7922	-0.1688	0	K.LWEIIDWDVVEK.R	(Ions score 54)
180 - 191	772.8390	1543.6634	1543.7922	-0.1288	0	K.LWEIIDWDVVEK.R	(Ions score 35)





```
LOCUS       YP_212180                193 aa            linear   BCT 01-MAY-2009
DEFINITION  superoxide dismutase [Fe] [Bacteroides fragilis NCTC 9343].
ACCESSION   YP_212180
VERSION     YP_212180.1  GI:60682036
DBLINK      Project:46
DBSOURCE    REFSEQ: accession NC_003228.3
KEYWORDS    .
SOURCE      Bacteroides fragilis NCTC 9343
  ORGANISM  Bacteroides fragilis NCTC 9343
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 193)
  AUTHORS   Cerdeno-Tarraga,A.M., Patrick,S., Crossman,L.C., Blakely,G.,
            Abratt,V., Lennard,N., Poxton,I., Duerden,B., Harris,B.,
            Quail,M.A., Barron,A., Clark,L., Corton,C., Doggett,J.,
            Holden,M.T., Larke,N., Line,A., Lord,A., Norbertczak,H., Ormond,D.,
            Price,C., Rabbinowitsch,E., Woodward,J., Barrell,B. and Parkhill,J.
  TITLE     Extensive DNA inversions in the B. fragilis genome control variable
            gene expression
  JOURNAL   Science 307 (5714), 1463-1465 (2005)
   PUBMED   15746427
REFERENCE   2  (residues 1 to 193)
  AUTHORS   Cerdeno-Tarraga,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2004) Pathogen Sequencing Unit, Sanger Institute,
            Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United
            Kingdom
REFERENCE   3  (residues 1 to 193)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from CAH08256.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..193
                     /organism="Bacteroides fragilis NCTC 9343"
                     /strain="ATCC 25285; NCTC 9343"
                     /db_xref="ATCC:25285"
                     /db_xref="taxon:272559"
     Protein         1..193
                     /product="superoxide dismutase [Fe]"
                     /EC_number="1.15.1.1"
                     /calculated_mol_wt=21640
     Region          1..193
                     /region_name="SodA"
                     /note="Superoxide dismutase [Inorganic ion transport and
                     metabolism]; COG0605"
                     /db_xref="CDD:30950"
     Region          2..83
                     /region_name="Sod_Fe_N"
                     /note="Iron/manganese superoxide dismutases, alpha-hairpin
                     domain; pfam00081"
                     /db_xref="CDD:109149"
     Region          90..192
                     /region_name="Sod_Fe_C"
                     /note="Iron/manganese superoxide dismutases, C-terminal
                     domain; pfam02777"
                     /db_xref="CDD:111650"
     CDS             1..193
                     /gene="sodB"
                     /locus_tag="BF2556"
                     /gene_synonym="sod"
                     /coded_by="NC_003228.3:2982352..2982933"
                     /note="Similar to Bacteroides fragilis superoxide
                     dismutase [Fe] SodB or Sod SWALL:SODF_BACFR (SWALL:P53638)
                     (193 aa) fasta scores: E(): 1.2e-77, 99.48% id in 193 aa,
                     and to Escherichia coli, Escherichia coli O6, Escherichia
                     coli O157:H7, and Shigella flexneri superoxide dismutase
                     [Fe] SodB or B1656 or C2050 or Z2678 or ECS2365 or SF1684
                     or S1816 SWALL:SODF_ECOLI (SWALL:P09157) (192 aa) fasta
                     scores: E(): 1.6e-37, 51.04% id in 192 aa"
                     /transl_table=11
                     /db_xref="GeneID:3287245"
```

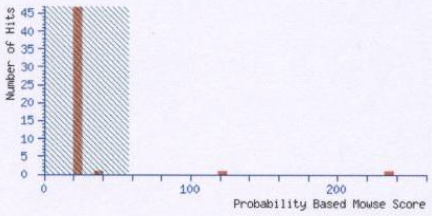
Mascot: http://www.matrixscience.com/

Mascot Search Results

```
User            : Lakshmy Manickan
Email           : lakshmy.manickan@unn.ac.uk
Search title    : SSP 7401
MS data file    : D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP 7401 (10)_RN15_01_1596.d\SSP 7401 (10)_RN15_01_1596.mgf
Database        : NCBInr 20100102 (10272453 sequences; 3505279183 residues)
Taxonomy        : Bacteria (Eubacteria) (5690016 sequences)
Timestamp       : 6 Jan 2010 at 16:24:05 GMT
Protein hits    : gi|53711978 ribosome recycling factor [Bacteroides fragilis YCH46]
                  gi|29347659 ribosome recycling factor [Bacteroides thetaiotaomicron VPI-5482]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 57 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As [Peptide Summary] Help

Significance threshold p< 0.05 Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ Ions score or expect cut-off 0 Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ Sort unassigned [Decreasing Score] Require bold red ☐

Select All | Select None | Search Selected ☐ **Error tolerant**

1. gi|53711978 **Mass:** 20809 **Score:** 235 **Queries matched:** 5 **emPAI:** 0.82
ribosome recycling factor [Bacteroides fragilis YCH46]
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	148	542.2400	1623.6982	1623.8443	-0.1461	1	37	4.9	1	R.SITIKPWDKSMFR.V + Oxidation (M)
☑	223	695.9640	2084.8702	2085.0412	-0.1710	0	(44)	0.82	1	K.AIIDSDLGIMPENNGEIIR.I + Oxidation (M)
☑	224	1043.9710	2085.9274	2085.0412	0.8863	0	86	6.3e-05	1	K.AIIDSDLGIMPENNGEIIR.I + Oxidation (M)
☑	241	1133.4620	2264.9094	2265.0947	-0.1853	0	75	0.00051	1	R.VDSYGSMVPISNVAAVTTPDAR.S + Oxidation (M)
☑	250	978.4010	2932.1812	2932.4964	-0.3153	1	37	3	1	R.LLDGIRVDSYGSMVPISNVAAVTTPDAR.S + Oxidation (M)

Proteins matching the same set of peptides:
gi|153808675 **Mass:** 20841 **Score:** 235 **Queries matched:** 5
hypothetical protein BACCAC_02974 [Bacteroides caccae ATCC 43185]
gi|265765322 **Mass:** 20823 **Score:** 235 **Queries matched:** 5
ribosome recycling factor [Bacteroides sp. 2_1_16]

---

2. gi|29347659 **Mass:** 20827 **Score:** 117 **Queries matched:** 3 **emPAI:** 0.16
ribosome recycling factor [Bacteroides thetaiotaomicron VPI-5482]
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
148	542.2400	1623.6982	1623.8443	-0.1461	1	37	4.9	1	R.SITIKPWDKSMFR.A + Oxidation (M)
241	1133.4620	2264.9094	2265.0947	-0.1853	0	53	0.082	2	R.VDSYGSMVPISNVAALSTPDAR.S + Oxidation (M)
250	978.4010	2932.1812	2932.4964	-0.3152	1	26	32	3	R.LLDGIRVDSYGSMVPISNVAALSTPDAR.S + Oxidation (M)

Peptide matches not assigned to protein hits: (no details means no match)

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	206	672.1310	2013.3712	2013.1016	0.2696	1	35	5.6	1	MPISSAPIRLPMRPIFR + 2 Oxidation (M)
☑	150	547.6960	1640.0662	1639.9257	0.1405	1	30	22	1	TVLTGDTLRALPDLR
☑	138	771.4460	2311.3162	2311.2172	0.0989	1	29	65	1	REGFVVTGTEAATPPETVPVVR
☑	115	657.2930	1312.5714	1311.6567	0.9148	0	28	44	1	LELGLTMGGFMK + Oxidation (M)
☑	246	883.6690	2647.9852	2647.0551	0.9301	0	27	28	1	NHDGGADVDELGYSCHPSAEDVMR + Oxidation (M)
☑	133	743.2970	2226.8692	2227.9435	-1.0744	0	27	93	1	QTDPVAAEAVVAECCNMAFK + Oxidation (M)
☑	193	658.3850	1972.1332	1972.0047	0.1284	1	26	57	1	LDEINLAMDRLADGSAIR
☑	162	901.3710	1800.7274	1800.9370	-0.2096	0	26	53	1	VETQTAAEVGGVIWVSR
☑	129	721.5120	2161.5142	2162.0439	-0.5297	2	26	1e+02	1	VHMEHELRGAWERDELR
☑	101	576.9530	1727.8372	1728.0371	-0.2000	1	26	1.2e+02	1	KTVFAALMLAAIPALAK

Mascot Search Results

Protein View

Match to: **gi|53711978** Score: **235**
ribosome recycling factor [Bacteroides fragilis YCH46]
Found in search of D:\Data\Lakshmy\PROTEOMICS 3- 151208\SSP 7401 (10)_RN15_01_1596.d\SSP 7401 (10)_RN15_01_1596.mgf

Nominal mass (M_r): **20809**; Calculated pI value: **5.86**
NCBI BLAST search of gi|53711978 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bacteroides fragilis YCH46
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|60680182 from Bacteroides fragilis NCTC 9343
gi|253563986 from Bacteroides sp. 3_2_5
gi|255007502 from Bacteroides fragilis 3_1_12
gi|60390242 from Bacteroides fragilis
gi|81316748 from Bacteroides fragilis NCTC 9343
gi|52214843 from Bacteroides fragilis YCH46
gi|60491616 from Bacteroides fragilis NCTC 9343
gi|251947762 from Bacteroides sp. 3_2_5

Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **32%**

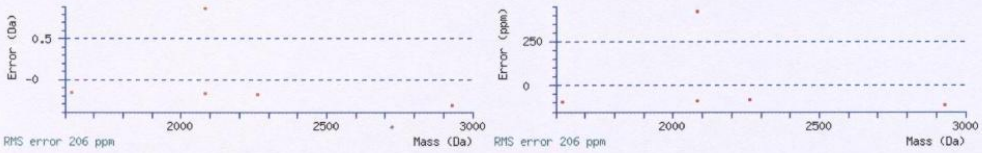
Matched peptides shown in Bold Red

```
  1 MVDVKTIIEE SQEKMDMAVM YLEEALAHIR AGKASTRLLD GIRVDSYGSM
 51 VPISNVAAVT TPDARSITIK PWDKSMFRVI EKAIIDSDLG IMPENNGEII
101 RIGIPPLTEE RRKQLAKQCK AEGETAKVSI RNARRDGIDA LKKAVKDGLA
151 EDEQKNAEAK LQKVHDKYIA KIEEMLAEKD KEIMTV
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
38 - 65	978.4010	2932.1812	2932.4964	-0.3153	1	R.LLDGIRVDSYGSMVPISNVAAVTTPDAR.S Oxidation (M) (Ions score 37)
44 - 65	1133.4620	2264.9094	2265.0947	-0.1853	0	R.VDSYGSMVPISNVAAVTTPDAR.S Oxidation (M) (Ions score 75)
66 - 78	542.2400	1623.6982	1623.8443	-0.1461	1	R.SITIKPWDKSMFR.V Oxidation (M) (Ions score 37)
83 - 101	695.9640	2084.8702	2085.0412	-0.1710	0	K.AIIDSDLGIMPENNGEIIR.I Oxidation (M) (Ions score 44)
83 - 101	1043.9710	2085.9274	2085.0412	0.8863	0	K.AIIDSDLGIMPENNGEIIR.I Oxidation (M) (Ions score 86)




```
LOCUS       YP_097970                186 aa            linear   BCT 26-APR-2009
DEFINITION  ribosome recycling factor [Bacteroides fragilis YCH46].
ACCESSION   YP_097970
VERSION     YP_097970.1  GI:53711978
DBLINK      Project:13067
DBSOURCE    REFSEQ: accession NC_006347.1
KEYWORDS    .
SOURCE      Bacteroides fragilis YCH46
  ORGANISM  Bacteroides fragilis YCH46
            Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (residues 1 to 186)
  AUTHORS   Kuwahara,T., Yamashita,A., Hirakawa,H., Nakayama,H., Toh,H.,
            Okada,N., Kuhara,S., Hattori,M., Hayashi,T. and Ohnishi,Y.
  TITLE     Genomic analysis of Bacteroides fragilis reveals extensive DNA
            inversions regulating cell surface adaptation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
   PUBMED   15466707
REFERENCE   2  (residues 1 to 186)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 186)
  AUTHORS   Hattori,M., Yamashita,A., Toh,H., Oshima,K. and Shiba,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2004) Kitasato Institute for Life Sciences,
            1-15-1, Kitasato, Sagamihara, Kanagawa 228-8555, Japan
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from BAD47436.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..186
                     /organism="Bacteroides fragilis YCH46"
                     /strain="YCH46"
                     /db_xref="taxon:295405"
     Protein         1..186
                     /product="ribosome recycling factor"
                     /calculated_mol_wt=20633
     Region          7..184
```

```
                /region_name="RRF"
                /note="Ribosome recycling factor (RRF). Ribosome recycling
                factor dissociates the posttermination complex, composed
                of the ribosome, deacylated tRNA, and mRNA, after
                termination of translation.  Thus ribosomes are 'recycled'
                and ready for another round of...; cd00520"
                /db_xref="CDD:29621"
Site            order(31..34,104..107)
                /site_type="other"
                /note="hinge region"
                /db_xref="CDD:29621"
CDS             1..186
                /gene="frr"
                /locus_tag="BF0689"
                /coded_by="complement(NC_006347.1:814916..815476)"
                /note="Rrf; Frr; ribosome-recycling factor; release factor
                4; RF4; recycles ribosomes upon translation termination
                along with release factor RF-3 and elongation factor EF-G;
                A GTPase-dependent process results in release of 50S from
                70S; inhibited by release factor RF-1; essential for
                viability; structurally similar to tRNAs"
                /transl_table=11
                /db_xref="GeneID:3081776"
```

Mascot: http://www.matrixscience.com/

G4 Blast analysis results for hypothetical proteins BF2494, BF1203 and BF0301 in *B. fragilis*

BLAST

Basic Local Alignment Search Tool

•

Edit and Resubmit Save Search Strategies Formatting options Download

gi|60491326|emb|CAH06074.1| putative transmembrane...

Results for: lcl|64607 gi|60491326|emb|CAH06074.1| putative transmembrane protein [Bacteroides fragilis NCTC 9343](406aa)
Your BLAST job specified more than one input sequence. This box lets you choose which input sequence to show BLAST results for.

Query ID
lcl|64607
lcl|64607
Description
gi|60491326|emb|CAH06074.1| putative transmembrane protein [Bacteroides fragilis NCTC 9343]
Molecule type
amino acid
Query Length
406

Database Name
nr
Description
All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program
BLASTP 2.2.22+ Citation

Reference
Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.
Reference - compositional score matrix adjustment
Stephen F. Altschul, John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schäffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.
Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Multiple alignment] NEW

Search Parameters

Search parameter name	Search parameter value
Program	blastp
Word size	3
Expect value	10
Hitlist size	100
Gapcosts	11,1
Matrix	BLOSUM62
Filter string	F
Genetic Code	1
Window Size	40
Threshold	11
Composition-based stats	2

Database

Database parameter name	Database parameter value
Posted date	Jan 20, 2010 4:14 PM
Number of letters	3,542,056,219
Number of sequences	10,381,779
Entrez query	none

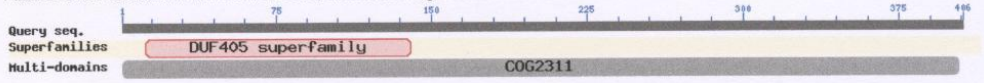
Karlin-Altschul statistics

Params	Ungapped	Gapped
Lambda	0.332433	0.267
K	0.144686	0.041
H	0.465429	0.14

Results Statistics

Results Statistics parameter name	Results Statistics parameter value
Length adjustment	139
Effective length of query	267
Effective length of database	2098988938
Effective search space	560430046446
Effective search space used	560430046446

Graphic Summary
Show Conserved Domains
Putative conserved domains have been detected, click on the image below for detailed results.



Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Sequences producing significant alignments:

(Click headers to sort columns)

YP_210036.1	putative transmembrane protein [Bacteroides fragilis NCTC 9343] >emb|CAH06074.1| putative transmembrane protein [Bacteroides fragilis NCTC 9343]	824	824	100%	0.0	G
ZP_04841755.1	conserved hypothetical protein [Bacteroides sp. 3_2_5] >ref|ZP_06093303.1| conserved hypothetical protein [Bacteroides sp. 2_1_16] >gb|EES88356.1| conserved hypothetical protein [Bacteroides sp. 3_2_5] >gb|EEZ25846.1| conserved hypothetical protein [Bacteroides sp. 2_1_16]	822	822	100%	0.0	
YP_097636.1	putative transport protein [Bacteroides fragilis YCH46] >dbj|BAD47102.1| putative transport protein [Bacteroides fragilis YCH46]	822	822	100%	0.0	G
YP_001818702.1	hypothetical protein Oter_1818 [Opitutus terrae PB90-1] >gb|ACB75102.1| protein of unknown function DUF418 [Opitutus terrae PB90-1]	261	261	97%	1e-67	G
ZP_05414587.1	putative membrane-associated protein [Bacteroides finegoldii DSM 17565] >gb|EEX46425.1| putative membrane-associated protein [Bacteroides finegoldii DSM 17565]	241	241	95%	8e-62	
NP_813370.1	putative transport protein [Bacteroides thetaiotaomicron VPI-5482] >ref|ZP_04847173.1| conserved hypothetical protein [Bacteroides sp. 1_1_6] >gb|AAO79564.1| conserved hypothetical protein, putative transport protein [Bacteroides thetaiotaomicron VPI-5482] >gb|EES68227.1| conserved hypothetical protein [Bacteroides sp. 1_1_6]	239	239	95%	4e-61	G
ZP_05757845.1	putative transport protein [Bacteroides sp. D2]	239	239	95%	5e-61	
ZP_02068252.1	hypothetical protein BACOVA_05266 [Bacteroides ovatus ATCC 8483] >gb|EDO09405.1| hypothetical protein BACOVA_05266 [Bacteroides ovatus ATCC 8483]	238	238	95%	1e-60	
ZP_04552009.1	conserved hypothetical protein [Bacteroides sp. 2_2_4] >gb|EEO55115.1| conserved hypothetical protein [Bacteroides sp. 2_2_4]	235	235	95%	7e-60	
ZP_05545092.1	conserved hypothetical protein [Parabacteroides sp. D13] >gb|EEU51838.1| conserved hypothetical protein [Parabacteroides sp. D13]	233	233	95%	3e-59	
YP_001302036.1	putative membrane-associated protein [Parabacteroides distasonis ATCC 8503] >gb|ABR42414.1| putative membrane-associated protein [Parabacteroides distasonis ATCC 8503]	233	233	95%	5e-59	G
ZP_05286207.1	putative membrane-associated protein [Bacteroides sp. 2_1_7]	231	231	95%	1e-58	
ZP_04846930.1	conserved hypothetical protein [Bacteroides sp. 1_1_6] >gb|EES69620.1| conserved hypothetical protein [Bacteroides sp. 1_1_6]	230	230	97%	2e-58	
ZP_06075273.1	conserved hypothetical protein [Bacteroides sp. 2_1_33B] >gb|EEY85242.1| conserved hypothetical protein [Bacteroides sp. 2_1_33B]	230	230	95%	2e-58	
NP_811692.1	putative membrane-associated protein [Bacteroides thetaiotaomicron VPI-5482] >gb|AAO77886.1| conserved hypothetical protein, putative membrane-associated protein [Bacteroides thetaiotaomicron VPI-5482]	228	228	97%	1e-57	G
ZP_04842518.1	conserved hypothetical protein [Bacteroides sp. 3_2_5] >ref|ZP_06094020.1| conserved hypothetical protein [Bacteroides sp. 2_1_16] >gb|EES86904.1| conserved hypothetical protein [Bacteroides sp. 3_2_5] >gb|EEZ25112.1| conserved hypothetical protein [Bacteroides sp. 2_1_16]	227	227	96%	2e-57	
YP_100550.1	putative membrane-associated protein [Bacteroides fragilis YCH46] >dbj|BAD50016.1| putative membrane-associated protein [Bacteroides fragilis YCH46]	226	226	96%	4e-57	G
YP_212723.1	hypothetical protein BF3110 [Bacteroides fragilis NCTC 9343] >emb|CAH08805.1| conserved hypothetical membrane protein [Bacteroides fragilis NCTC 9343]	226	226	96%	5e-57	G
ZP_03014090.1	hypothetical protein BACINT_01653 [Bacteroides intestinalis DSM 17393] >gb|EDV06564.1| hypothetical protein BACINT_01653 [Bacteroides intestinalis DSM 17393]	224	224	96%	1e-56	
ZP_03678199.1	hypothetical protein BACCELL_02542 [Bacteroides cellulosilyticus DSM 14838] >gb|EEF89824.1| hypothetical protein BACCELL_02542 [Bacteroides cellulosilyticus DSM 14838]	223	223	96%	3e-56	
ZP_05282878.1	hypothetical protein Bfra3_16553 [Bacteroides fragilis 3_1_12]	223	223	96%	4e-56	
ZP_03643854.1	hypothetical protein BACCOPRO_02228 [Bacteroides coprophilus DSM 18228] >gb|EEF76722.1| hypothetical protein BACCOPRO_02228 [Bacteroides coprophilus DSM 18228]	218	218	95%	1e-54	
YP_001299550.1	hypothetical protein BVU_2269 [Bacteroides vulgatus ATCC 8482] >gb|ABR39928.1| conserved hypothetical membrane protein [Bacteroides vulgatus ATCC 8482]	218	218	96%	1e-54	G
ZP_05255382.1	conserved hypothetical protein [Bacteroides sp. 4_3_47FAA] >gb|EET15774.1| conserved hypothetical protein [Bacteroides sp. 4_3_47FAA]	217	217	96%	3e-54	
ZP_04558306.1	conserved hypothetical protein [Bacteroides sp. D4] >ref|ZP_06087176.1| conserved hypothetical protein [Bacteroides sp. 3_1_33FAA] >gb|EEO44758.1| conserved hypothetical protein [Bacteroides sp. D4] >gb|EEZ23459.1| conserved hypothetical protein [Bacteroides sp. 3_1_33FAA]	216	216	96%	5e-54	
ZP_04540787.1	conserved hypothetical protein [Bacteroides sp. 9_1_42FAA] >gb|EEO61489.1| conserved hypothetical protein [Bacteroides sp. 9_1_42FAA]	215	215	96%	1e-53	
ZP_03302157.1	hypothetical protein BACDOR_03555 [Bacteroides dorei DSM 17855] >gb|EEB24135.1| hypothetical protein BACDOR_03555 [Bacteroides dorei DSM 17855]	213	213	96%	2e-53	

Alignments Select All Get selected sequences Distance tree of results Multiple alignment NEW

```
>ref|YP_210036.1| G putative transmembrane protein [Bacteroides fragilis NCTC 9343]
 emb|CAH06074.1| G putative transmembrane protein [Bacteroides fragilis NCTC 9343]
Length=406

 GENE ID: 3285991 BF0301 | putative transmembrane protein
[Bacteroides fragilis NCTC 9343] (10 or fewer PubMed links)

 Score =  824 bits (2129),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 406/406 (100%), Positives = 406/406 (100%), Gaps = 0/406 (0%)

Query  1    MTHTQTITPKKRINSIDALRGFALIGIMLLHCMERFDLTLAPVVESPFWQAIDTAVYDSL  60
            MTHTQTITPKKRINSIDALRGFALIGIMLLHCMERFDLTLAPVVESPFWQAIDTAVYDSL
Sbjct  1    MTHTQTITPKKRINSIDALRGFALIGIMLLHCMERFDLTLAPVVESPFWQAIDTAVYDSL  60

Query  61   YFLFSGKSYAMFSLLFGLSFFMQMESQAAKGVDFRGRFLWRLALLFLFGYINGLVYMGEF  120
            YFLFSGKSYAMFSLLFGLSFFMQMESQAAKGVDFRGRFLWRLALLFLFGYINGLVYMGEF
Sbjct  61   YFLFSGKSYAMFSLLFGLSFFMQMESQAAKGVDFRGRFLWRLALLFLFGYINGLVYMGEF  120

Query  121  FMVYAVLGVFLIPLYKVPTRWLLVLCVLLFLQIPAVISFVSLLSDNVANEPTAAAAYMDR  180
            FMVYAVLGVFLIPLYKVPTRWLLVLCVLLFLQIPAVISFVSLLSDNVANEPTAAAAYMDR
Sbjct  121  FMVYAVLGVFLIPLYKVPTRWLLVLCVLLFLQIPAVISFVSLLSDNVANEPTAAAAYMDR  180

Query  181  LFERAADVFINGSLMDVLSFNTFDGQSAKCLWVFNNFRYLQLLGLFIAGMLIGRQGIHKS  240
            LFERAADVFINGSLMDVLSFNTFDGQSAKCLWVFNNFRYLQLLGLFIAGMLIGRQGIHKS
Sbjct  181  LFERAADVFINGSLMDVLSFNTFDGQSAKCLWVFNNFRYLQLLGLFIAGMLIGRQGIHKS  240

Query  241  EEKMVKYSRLFLPYCLAFWAVFYAVAFLLPVWGVDGFALRVGQTLFKTYGNLGQMMVYFC  300
            EEKMVKYSRLFLPYCLAFWAVFYAVAFLLPVWGVDGFALRVGQTLFKTYGNLGQMMVYFC
Sbjct  241  EEKMVKYSRLFLPYCLAFWAVFYAVAFLLPVWGVDGFALRVGQTLFKTYGNLGQMMVYFC  300

Query  301  GFTLLYYRYKGQKVLDRIAPVGRMSVTNYMAQSIVGVSLFYGFGGNFAVEFNYLQSFLLG  360
            GFTLLYYRYKGQKVLDRIAPVGRMSVTNYMAQSIVGVSLFYGFGGNFAVEFNYLQSFLLG
Sbjct  301  GFTLLYYRYKGQKVLDRIAPVGRMSVTNYMAQSIVGVSLFYGFGGNFAVEFNYLQSFLLG  360

Query  361  AAFCVIQIAYSNWWIKRFYYGPMEWLWRSLTWFQVVPLSRRKASLG  406
            AAFCVIQIAYSNWWIKRFYYGPMEWLWRSLTWFQVVPLSRRKASLG
Sbjct  361  AAFCVIQIAYSNWWIKRFYYGPMEWLWRSLTWFQVVPLSRRKASLG  406


>ref|ZP_04841755.1|  conserved hypothetical protein [Bacteroides sp. 3_2_5]
 ref|ZP_06093303.1|  conserved hypothetical protein [Bacteroides sp. 2_1_16]
 gb|EES88356.1|  conserved hypothetical protein [Bacteroides sp. 3_2_5]
 gb|EEZ25846.1|  conserved hypothetical protein [Bacteroides sp. 2_1_16]
Length=412

 Score =  822 bits (2123),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 404/406 (99%), Positives = 405/406 (99%), Gaps = 0/406 (0%)

Query  1    MTHTQTITPKKRINSIDALRGFALIGIMLLHCMERFDLTLAPVVESPFWQAIDTAVYDSL  60
            MTHTQTITPKKRINSIDALRGFALIGIMLLHCMERFDLTLAPVVESPFWQAIDTAVYDSL
Sbjct  7    MTHTQTITPKKRINSIDALRGFALIGIMLLHCMERFDLTLAPVVESPFWQAIDTAVYDSL  66

Query  61   YFLFSGKSYAMFSLLFGLSFFMQMESQAAKGVDFRGRFLWRLALLFLFGYINGLVYMGEF  120
            YFLFSGKSYAMFSLLFGLSFFMQMESQAAKGVDFRGRFLWRLALLFLFGYINGLVYMGEF
Sbjct  67   YFLFSGKSYAMFSLLFGLSFFMQMESQAAKGVDFRGRFLWRLALLFLFGYINGLVYMGEF  126

Query  121  FMVYAVLGVFLIPLYKVPTRWLLVLCVLLFLQIPAVISFVSLLSDNVANEPTAAAAYMDR  180
            FMVYAVLGVFLIPLYKV TRWLLVLC+LLFLQIPAVISFVSLLSDNVANEPTAAAAYMDR
Sbjct  127  FMVYAVLGVFLIPLYKVSTRWLLVLCILLFLQIPAVISFVSLLSDNVANEPTAAAAYMDR  186

Query  181  LFERAADVFINGSLMDVLSFNTFDGQSAKCLWVFNNFRYLQLLGLFIAGMLIGRQGIHKS  240
            LFERAADVFINGSLMDVLSFNTFDGQSAKCLWVFNNFRYLQLLGLFIAGMLIGRQGIHKS
Sbjct  187  LFERAADVFINGSLMDVLSFNTFDGQSAKCLWVFNNFRYLQLLGLFIAGMLIGRQGIHKS  246

Query  241  EEKMVKYSRLFLPYCLAFWAVFYAVAFLLPVWGVDGFALRVGQTLFKTYGNLGQMMVYFC  300
            EEKMVKYSRLFLPYCLAFWAVFYAVAFLLPVWGVDGFALRVGQTLFKTYGNLGQMMVYFC
Sbjct  247  EEKMVKYSRLFLPYCLAFWAVFYAVAFLLPVWGVDGFALRVGQTLFKTYGNLGQMMVYFC  306

Query  301  GFTLLYYRYKGQKVLDRIAPVGRMSVTNYMAQSIVGVSLFYGFGGNFAVEFNYLQSFLLG  360
            GFTLLYYRYKGQKVLDRIAPVGRMSVTNYMAQSIVGVSLFYGFGGNFAVEFNYLQSFLLG
Sbjct  307  GFTLLYYRYKGQKVLDRIAPVGRMSVTNYMAQSIVGVSLFYGFGGNFAVEFNYLQSFLLG  366

Query  361  AAFCVIQIAYSNWWIKRFYYGPMEWLWRSLTWFQVVPLSRRKASLG  406
            AAFCVIQIAYSNWWIKRFYYGPMEWLWRSLTWFQVVPLSRRKASLG
Sbjct  367  AAFCVIQIAYSNWWIKRFYYGPMEWLWRSLTWFQVVPLSRRKASLG  412


>ref|YP_097636.1| G putative transport protein [Bacteroides fragilis YCH46]
 dbj|BAD47102.1| G putative transport protein [Bacteroides fragilis YCH46]
Length=406

 GENE ID: 3081573 BF0353 | putative transport protein
[Bacteroides fragilis YCH46] (10 or fewer PubMed links)

 Score =  822 bits (2123),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 405/406 (99%), Positives = 405/406 (99%), Gaps = 0/406 (0%)

Query  1    MTHTQTITPKKRINSIDALRGFALIGIMLLHCMERFDLTLAPVVESPFWQAIDTAVYDSL  60
            MTHTQTITPKKRINSIDALRGFALIGIMLLHCMERFDLTLAPVVESPFWQAIDTAVYDSL
Sbjct  1    MTHTQTITPKKRINSIDALRGFALIGIMLLHCMERFDLTLAPVVESPFWQAIDTAVYDSL  60

Query  61   YFLFSGKSYAMFSLLFGLSFFMQMESQAAKGVDFRGRFLWRLALLFLFGYINGLVYMGEF  120
            YFLFSGKSYAMFSLLFGLSFFMQMESQAAKGVDFRGRFLWRLALLFLFGYINGLVYMGEF
Sbjct  61   YFLFSGKSYAMFSLLFGLSFFMQMESQAAKGVDFRGRFLWRLALLFLFGYINGLVYMGEF  120

Query  121  FMVYAVLGVFLIPLYKVPTRWLLVLCVLLFLQIPAVISFVSLLSDNVANEPTAAAAYMDR  180
            FMVYAVLGVFLIPLYKV TRWLLVLCVLLFLQIPAVISFVSLLSDNVANEPTAAAAYMDR
Sbjct  121  FMVYAVLGVFLIPLYKVSTRWLLVLCVLLFLQIPAVISFVSLLSDNVANEPTAAAAYMDR  180

Query  181  LFERAADVFINGSLMDVLSFNTFDGQSAKCLWVFNNFRYLQLLGLFIAGMLIGRQGIHKS  240
            LFERAADVFINGSLMDVLSFNTFDGQSAKCLWVFNNFRYLQLLGLFIAGMLIGRQGIHKS
Sbjct  181  LFERAADVFINGSLMDVLSFNTFDGQSAKCLWVFNNFRYLQLLGLFIAGMLIGRQGIHKS  240

Query  241  EEKMVKYSRLFLPYCLAFWAVFYAVAFLLPVWGVDGFALRVGQTLFKTYGNLGQMMVYFC  300
            EEKMVKYSRLFLPYCLAFWAVFYAVAFLLPVWGVDGFALRVGQTLFKTYGNLGQMMVYFC
Sbjct  241  EEKMVKYSRLFLPYCLAFWAVFYAVAFLLPVWGVDGFALRVGQTLFKTYGNLGQMMVYFC  300

Query  301  GFTLLYYRYKGQKVLDRIAPVGRMSVTNYMAQSIVGVSLFYGFGGNFAVEFNYLQSFLLG  360
            GFTLLYYRYKGQKVLDRIAPVGRMSVTNYMAQSIVGVSLFYGFGGNFAVEFNYLQSFLLG
```

BLAST

Basic Local Alignment Search Tool

•

Edit and Resubmit Save Search Strategies Formatting options Download

gi|60492160|emb|CAH06923.1| putative anti-sigma...

Results for: lcl|86207 gi|60492160|emb|CAH06923.1| putative anti-sigma factor [Bacteroides fragilis NCTC 9343](309aa)
Your BLAST job specified more than one input sequence. This box lets you choose which input sequence to show BLAST results for.

Query ID
lcl|86207
lcl|86207
Description
gi|60492160|emb|CAH06923.1| putative anti-sigma factor [Bacteroides fragilis NCTC 9343]
Molecule type
amino acid
Query Length
309

Database Name
nr
Description
All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program
BLASTP 2.2.22+ Citation

Reference
Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.
Reference - compositional score matrix adjustment
Stephen F. Altschul, John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schäffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.
Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Multiple alignment] NEW

Search Parameters

Search parameter name	Search parameter value
Program	blastp
Word size	3
Expect value	10
Hitlist size	100
Gapcosts	11,1
Matrix	BLOSUM62
Filter string	F
Genetic Code	1
Window Size	40
Threshold	11
Composition-based stats	2

Database

Database parameter name	Database parameter value
Posted date	Jan 20, 2010 4:14 PM
Number of letters	3,542,056,219
Number of sequences	10,381,779
Entrez query	none

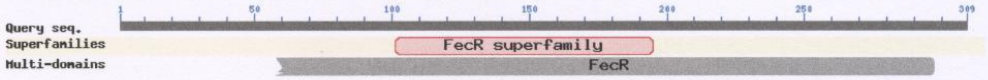
Karlin-Altschul statistics

Params	Ungapped	Gapped
Lambda	0.319277	0.267
K	0.135008	0.041
H	0.381721	0.14

Results Statistics

Results Statistics parameter name	Results Statistics parameter value
Length adjustment	136
Effective length of query	173
Effective length of database	2130134275
Effective search space	368513229575
Effective search space used	368513229575

Graphic Summary
Show Conserved Domains
Putative conserved domains have been detected, click on the image below for detailed results.



Alignments Select All Get selected sequences Distance tree of results Multiple alignment NEW

```
>ref|YP_098536.1| G putative anti-sigma factor [Bacteroides fragilis YCH46]
 ref|ZP_06091266.1|  conserved hypothetical protein [Bacteroides sp. 2_1_16]
 dbj|BAD48002.1| G putative anti-sigma factor [Bacteroides fragilis YCH46]
 gb|EEZ26652.1|  conserved hypothetical protein [Bacteroides sp. 2_1_16]
Length=321

 GENE ID: 3082788 BF1252 | putative anti-sigma factor
[Bacteroides fragilis YCH46] (10 or fewer PubMed links)

 Score =  635 bits (1638),  Expect = 2e-180, Method: Compositional matrix adjust.
 Identities = 309/309 (100%), Positives = 309/309 (100%), Gaps = 0/309 (0%)

Query  1    MTCEEKKDLFDRIESDEALKKEFLRMQNVVALTQILSRQDDSETSRKGKQHFMQLLFRKR  60
            MTCEEKKDLFDRIESDEALKKEFLRMQNVVALTQILSRQDDSETSRKGKQHFMQLLFRKR
Sbjct  13   MTCEEKKDLFDRIESDEALKKEFLRMQNVVALTQILSRQDDSETSRKGKQHFMQLLFRKR  72

Query  61   LKRAITVSLKYAAVFAVLVVGTFYTAKLYLSEEFGKSYTIVTAPKGQRVKIELPDGTIAW  120
            LKRAITVSLKYAAVFAVLVVGTFYTAKLYLSEEFGKSYTIVTAPKGQRVKIELPDGTIAW
Sbjct  73   LKRAITVSLKYAAVFAVLVVGTFYTAKLYLSEEFGKSYTIVTAPKGQRVKIELPDGTIAW  132

Query  121  LSPCSRLRFAASFNETDRKIELDGATYFDVAKNPEKPFVVSAKGYRIRVLGTKFNISAYK  180
            LSPCSRLRFAASFNETDRKIELDGATYFDVAKNPEKPFVVSAKGYRIRVLGTKFNISAYK
Sbjct  133  LSPCSRLRFAASFNETDRKIELDGATYFDVAKNPEKPFVVSAKGYRIRVLGTKFNISAYK  192

Query  181  NSKEFETDLVEGCVHIYDPADIRNEVFLQPKEKAVLWGDRLMKRESDFDNEEYLKNGVVS  240
            NSKEFETDLVEGCVHIYDPADIRNEVFLQPKEKAVLWGDRLMKRESDFDNEEYLKNGVVS
Sbjct  193  NSKEFETDLVEGCVHIYDPADIRNEVFLQPKEKAVLWGDRLMKRESDFDNEEYLKNGVVS  252

Query  241  FLSEPFGRVLNSVALWNDVNIKIERSVNATQRISGKFRQSDSLESILKALQGAMPFKYKI  300
            FLSEPFGRVLNSVALWNDVNIKIERSVNATQRISGKFRQSDSLESILKALQGAMPFKYKI
Sbjct  253  FLSEPFGRVLNSVALWNDVNIKIERSVNATQRISGKFRQSDSLESILKALQGAMPFKYKI  312

Query  301  VSEEEIIIY  309
            VSEEEIIIY
Sbjct  313  VSEEEIIIY  321

>ref|YP_210870.1| G putative anti-sigma factor [Bacteroides fragilis NCTC 9343]
 emb|CAH06923.1| G putative anti-sigma factor [Bacteroides fragilis NCTC 9343]
Length=309

 GENE ID: 3288859 BF1203 | putative anti-sigma factor
[Bacteroides fragilis NCTC 9343] (10 or fewer PubMed links)

 Score =  635 bits (1637),  Expect = 3e-180, Method: Compositional matrix adjust.
 Identities = 309/309 (100%), Positives = 309/309 (100%), Gaps = 0/309 (0%)

Query  1    MTCEEKKDLFDRIESDEALKKEFLRMQNVVALTQILSRQDDSETSRKGKQHFMQLLFRKR  60
            MTCEEKKDLFDRIESDEALKKEFLRMQNVVALTQILSRQDDSETSRKGKQHFMQLLFRKR
Sbjct  1    MTCEEKKDLFDRIESDEALKKEFLRMQNVVALTQILSRQDDSETSRKGKQHFMQLLFRKR  60

Query  61   LKRAITVSLKYAAVFAVLVVGTFYTAKLYLSEEFGKSYTIVTAPKGQRVKIELPDGTIAW  120
            LKRAITVSLKYAAVFAVLVVGTFYTAKLYLSEEFGKSYTIVTAPKGQRVKIELPDGTIAW
Sbjct  61   LKRAITVSLKYAAVFAVLVVGTFYTAKLYLSEEFGKSYTIVTAPKGQRVKIELPDGTIAW  120

Query  121  LSPCSRLRFAASFNETDRKIELDGATYFDVAKNPEKPFVVSAKGYRIRVLGTKFNISAYK  180
            LSPCSRLRFAASFNETDRKIELDGATYFDVAKNPEKPFVVSAKGYRIRVLGTKFNISAYK
Sbjct  121  LSPCSRLRFAASFNETDRKIELDGATYFDVAKNPEKPFVVSAKGYRIRVLGTKFNISAYK  180
                                                  *
Query  181  NSKEFETDLVEGCVHIYDPADIRNEVFLQPKEKAVLWGDRLMKRESDFDNEEYLKNGVVS  240
            NSKEFETDLVEGCVHIYDPADIRNEVFLQPKEKAVLWGDRLMKRESDFDNEEYLKNGVVS
Sbjct  181  NSKEFETDLVEGCVHIYDPADIRNEVFLQPKEKAVLWGDRLMKRESDFDNEEYLKNGVVS  240

Query  241  FLSEPFGRVLNSVALWNDVNIKIERSVNATQRISGKFRQSDSLESILKALQGAMPFKYKI  300
            FLSEPFGRVLNSVALWNDVNIKIERSVNATQRISGKFRQSDSLESILKALQGAMPFKYKI
Sbjct  241  FLSEPFGRVLNSVALWNDVNIKIERSVNATQRISGKFRQSDSLESILKALQGAMPFKYKI  300

Query  301  VSEEEIIIY  309
            VSEEEIIIY
Sbjct  301  VSEEEIIIY  309

>ref|ZP_04840883.1|  conserved hypothetical protein [Bacteroides sp. 3_2_5]
 gb|EESB7484.1|  conserved hypothetical protein [Bacteroides sp. 3_2_5]
Length=321

 Score =  631 bits (1627),  Expect = 3e-179, Method: Compositional matrix adjust.
 Identities = 308/309 (99%), Positives = 308/309 (99%), Gaps = 0/309 (0%)

Query  1    MTCEEKKDLFDRIESDEALKKEFLRMQNVVALTQILSRQDDSETSRKGKQHFMQLLFRKR  60
            MTCEEKKDLFDRIESDEALKKEFLRMQNVVALTQILS QDDSETSRKGKQHFMQLLFRKR
Sbjct  13   MTCEEKKDLFDRIESDEALKKEFLRMQNVVALTQILSCQDDSETSRKGKQHFMQLLFRKR  72

Query  61   LKRAITVSLKYAAVFAVLVVGTFYTAKLYLSEEFGKSYTIVTAPKGQRVKIELPDGTIAW  120
            LKRAITVSLKYAAVFAVLVVGTFYTAKLYLSEEFGKSYTIVTAPKGQRVKIELPDGTIAW
Sbjct  73   LKRAITVSLKYAAVFAVLVVGTFYTAKLYLSEEFGKSYTIVTAPKGQRVKIELPDGTIAW  132

Query  121  LSPCSRLRFAASFNETDRKIELDGATYFDVAKNPEKPFVVSAKGYRIRVLGTKFNISAYK  180
            LSPCSRLRFAASFNETDRKIELDGATYFDVAKNPEKPFVVSAKGYRIRVLGTKFNISAYK
Sbjct  133  LSPCSRLRFAASFNETDRKIELDGATYFDVAKNPEKPFVVSAKGYRIRVLGTKFNISAYK  192

Query  181  NSKEFETDLVEGCVHIYDPADIRNEVFLQPKEKAVLWGDRLMKRESDFDNEEYLKNGVVS  240
            NSKEFETDLVEGCVHIYDPADIRNEVFLQPKEKAVLWGDRLMKRESDFDNEEYLKNGVVS
Sbjct  193  NSKEFETDLVEGCVHIYDPADIRNEVFLQPKEKAVLWGDRLMKRESDFDNEEYLKNGVVS  252

Query  241  FLSEPFGRVLNSVALWNDVNIKIERSVNATQRISGKFRQSDSLESILKALQGAMPFKYKI  300
            FLSEPFGRVLNSVALWNDVNIKIERSVNATQRISGKFRQSDSLESILKALQGAMPFKYKI
Sbjct  253  FLSEPFGRVLNSVALWNDVNIKIERSVNATQRISGKFRQSDSLESILKALQGAMPFKYKI  312

Query  301  VSEEEIIIY  309
            VSEEEIIIY
Sbjct  313  VSEEEIIIY  321

>ref|ZP_05280128.1|  putative anti-sigma factor [Bacteroides fragilis 3_1_12]
Length=321

 Score =  586 bits (1511),  Expect = 1e-165, Method: Compositional matrix adjust.
 Identities = 279/309 (90%), Positives = 293/309 (94%), Gaps = 0/309 (0%)
```

Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Sequences producing significant alignments:

(Click headers to sort columns)

YP_098536.1	putative anti-sigma factor [Bacteroides fragilis YCH46] >ref|ZP_06091266.1| conserved hypothetical protein [Bacteroides sp. 2_1_16] >dbj|BAD48002.1| putative anti-sigma factor [Bacteroides fragilis YCH46] >gb|EEZ26652.1| conserved hypothetical protein [Bacteroides sp. 2_1_16]	635	635	100%	2e-180	G
YP_210870.1	putative anti-sigma factor [Bacteroides fragilis NCTC 9343] >emb|CAH06923.1| putative anti-sigma factor [Bacteroides fragilis NCTC 9343]	635	635	100%	3e-180	G
ZP_04840883.1	conserved hypothetical protein [Bacteroides sp. 3_2_5] >gb|EES87484.1| conserved hypothetical protein [Bacteroides sp. 3_2_5]	631	631	100%	3e-179	
ZP_05280128.1	putative anti-sigma factor [Bacteroides fragilis 3_1_12]	586	586	100%	1e-165	
ZP_05283939.1	hypothetical protein Bfra3_21925 [Bacteroides fragilis 3_1_12]	252	252	100%	3e-65	
YP_101246.1	putative anti-sigma factor [Bacteroides fragilis YCH46] >ref|YP_213335.1| hypothetical protein BF3745 [Bacteroides fragilis NCTC 9343] >ref|ZP_04844595.1| conserved hypothetical protein [Bacteroides sp. 3_2_5] >dbj|BAD50712.1| putative anti-sigma factor [Bacteroides fragilis YCH46] >emb|CAH09426.1| putative membrane protein [Bacteroides fragilis NCTC 9343] >gb|EES84495.1| conserved hypothetical protein [Bacteroides sp. 3_2_5]	250	250	100%	1e-64	G
ZP_06095273.1	conserved hypothetical protein [Bacteroides sp. 2_1_16] >gb|EEZ23941.1| conserved hypothetical protein [Bacteroides sp. 2_1_16]	250	250	100%	1e-64	
NP_811110.1	putative anti-sigma factor [Bacteroides thetaiotaomicron VPI-5482] >gb|AAO77304.1| putative anti-sigma factor [Bacteroides thetaiotaomicron VPI-5482]	230	230	99%	2e-58	G
ZP_03476000.1	hypothetical protein PRABACTJOHN_01664 [Parabacteroides johnsonii DSM 18315] >gb|EEC96931.1| hypothetical protein PRABACTJOHN_01664 [Parabacteroides johnsonii DSM 18315]	180	180	98%	2e-43	
ZP_03677961.1	hypothetical protein BACCELL_02300 [Bacteroides cellulosilyticus DSM 14838] >gb|EEF90053.1| hypothetical protein BACCELL_02300 [Bacteroides cellulosilyticus DSM 14838]	176	176	96%	4e-42	
ZP_02034096.1	hypothetical protein PARMER_04138 [Parabacteroides merdae ATCC 43184] >gb|EDN84686.1| hypothetical protein PARMER_04138 [Parabacteroides merdae ATCC 43184]	168	168	99%	8e-40	
ZP_05414155.1	putative anti-sigma factor [Bacteroides finegoldii DSM 17565] >gb|EEX46802.1| putative anti-sigma factor [Bacteroides finegoldii DSM 17565]	161	161	99%	7e-38	
ZP_04544887.1	conserved hypothetical protein [Bacteroides sp. D1] >ref|ZP_06084784.1| conserved hypothetical protein [Bacteroides sp. 2_1_22] >gb|EEO51361.1| conserved hypothetical protein [Bacteroides sp. D1] >gb|EEZ02882.1| conserved hypothetical protein [Bacteroides sp. 2_1_22]	161	161	99%	8e-38	
ZP_05758326.1	putative anti-sigma factor [Bacteroides sp. D2]	159	159	99%	3e-37	
ZP_03677876.1	hypothetical protein BACCELL_02215 [Bacteroides cellulosilyticus DSM 14838] >gb|EEF90135.1| hypothetical protein BACCELL_02215 [Bacteroides cellulosilyticus DSM 14838]	154	154	95%	1e-35	
ZP_03678231.1	hypothetical protein BACCELL_02574 [Bacteroides cellulosilyticus DSM 14838] >gb|EEF89800.1| hypothetical protein BACCELL_02574 [Bacteroides cellulosilyticus DSM 14838]	154	154	99%	2e-35	
ZP_02435836.1	hypothetical protein BACSTE_02087 [Bacteroides stercoris ATCC 43183] >gb|EDS14404.1| hypothetical protein BACSTE_02087 [Bacteroides stercoris ATCC 43183]	153	153	99%	2e-35	
ZP_02068886.1	hypothetical protein BACUNI_00286 [Bacteroides uniformis ATCC 8492] >gb|EDO56000.1| hypothetical protein BACUNI_00286 [Bacteroides uniformis ATCC 8492]	152	152	99%	3e-35	
ZP_03014123.1	hypothetical protein BACINT_01686 [Bacteroides intestinalis DSM 17393] >gb|EDV06597.1| hypothetical protein BACINT_01686 [Bacteroides intestinalis DSM 17393]	152	152	99%	3e-35	
ZP_03299904.1	hypothetical protein BACDOR_01271 [Bacteroides dorei DSM 17855] >ref|ZP_04554581.1| conserved hypothetical protein [Bacteroides sp. D4] >gb|EEB26249.1| hypothetical protein BACDOR_01271 [Bacteroides dorei DSM 17855] >gb|EEO47637.1| conserved hypothetical protein [Bacteroides sp. D4]	152	152	87%	5e-35	
ZP_03477618.1	hypothetical protein PRABACTJOHN_03305 [Parabacteroides johnsonii DSM 18315] >gb|EEC95310.1| hypothetical protein PRABACTJOHN_03305 [Parabacteroides johnsonii DSM 18315]	152	152	96%	5e-35	
YP_001304665.1	putative anti-sigma factor [Parabacteroides distasonis ATCC 8503] >gb|ABR45043.1| putative anti-sigma factor [Parabacteroides distasonis ATCC 8503]	152	152	95%	5e-35	G
YP_001297833.1	putative anti-sigma factor [Bacteroides vulgatus ATCC 8482] >gb|ABR38211.1| putative anti-sigma factor [Bacteroides vulgatus ATCC 8482]	152	152	87%	6e-35	G
ZP_05255078.1	conserved hypothetical protein [Bacteroides sp. 4_3_47FAA] >gb|EET15470.1| conserved hypothetical protein [Bacteroides sp. 4_3_47FAA]	151	151	87%	8e-35	
YP_101600.1	putative anti-sigma factor [Bacteroides fragilis YCH46] >ref|YP_213696.1| putative anti-sigma factor [Bacteroides fragilis NCTC 9343] >ref|ZP_04844033.1| conserved hypothetical protein [Bacteroides sp. 3_2_5] >ref|ZP_06095115.1| conserved hypothetical protein [Bacteroides sp. 2_1_16] >dbj|BAD51066.1| putative anti-sigma factor [Bacteroides fragilis YCH46] >emb|CAH09804.1| putative anti-sigma factor [Bacteroides fragilis NCTC 9343] >gb|EES85227.1| conserved hypothetical protein [Bacteroides sp. 3_2_5] >gb|EEZ24266.1| conserved hypothetical protein [Bacteroides	150	150	81%	1e-34	G

BLAST

Basic Local Alignment Search Tool

•

Edit and Resubmit Save Search Strategies Formatting options Download

gi|60681974|ref|YP_212118.1| hypothetical...

Results for: lcl|73919 gi|60681974|ref|YP_212118.1| hypothetical protein BF2494 [Bacteroides fragilis NCTC 9343](396aa)
Your BLAST job specified more than one input sequence. This box lets you choose which input sequence to show BLAST results for.

Query ID
lcl|73919
lcl|73919
Description
gi|60681974|ref|YP_212118.1| hypothetical protein BF2494 [Bacteroides fragilis NCTC 9343]
Molecule type
amino acid
Query Length
396

Database Name
nr
Description
All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program
BLASTP 2.2.22+ Citation

Reference
Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.
Reference - compositional score matrix adjustment
Stephen F. Altschul, John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schäffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.
Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Related Structures] [Multiple alignment] NEW

Search Parameters

Search parameter name	Search parameter value
Program	blastp
Word size	3
Expect value	10
Hitlist size	100
Gapcosts	11,1
Matrix	BLOSUM62
Filter string	F
Genetic Code	1
Window Size	40
Threshold	11
Composition-based stats	2

Database

Database parameter name	Database parameter value
Posted date	Jan 20, 2010 4:14 PM
Number of letters	3,542,056,219
Number of sequences	10,381,779
Entrez query	none

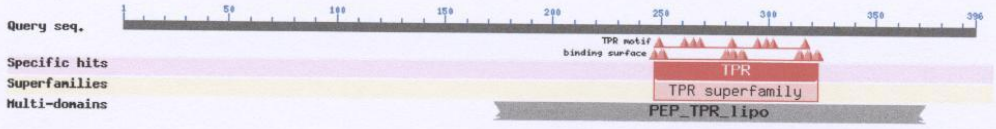
Karlin-Altschul statistics

Params	Ungapped	Gapped
Lambda	0.314287	0.267
K	0.13096	0.041
H	0.371554	0.14

Results Statistics

Results Statistics parameter name	Results Statistics parameter value
Length adjustment	139
Effective length of query	257
Effective length of database	2098988938
Effective search space	539440157066
Effective search space used	539440157066

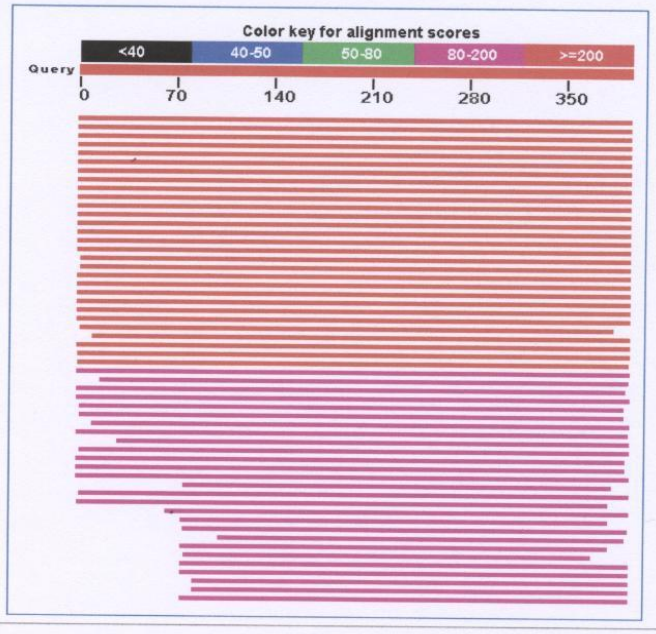
Graphic Summary
Show Conserved Domains
Putative conserved domains have been detected, click on the image below for detailed results.



Distribution of 140 Blast Hits on the Query Sequence

[?]

An overview of the database sequences aligned to the query sequence is shown. The score of each alignment is indicated by one of five different colors, which divides the range of scores into five groups. Multiple alignments on the same database sequence are connected by a striped line. Mousing over a hit sequence causes the definition and score to be shown in the window at the top, clicking on a hit sequence takes the user to the associated alignments. New: This graphic is an overview of database sequences aligned to the query sequence. Alignments are color-coded by score, within one of five score ranges. Multiple alignments on the same database sequence are connected by a dashed line. Mousing over an alignment shows the alignment definition and score in the box at the top. Clicking an alignment displays the alignment detail.



Descriptions

Sequences producing significant alignments:	Score (Bits)	E Value	
ref|YP_099695.1| TPR repeat-containing protein [Bacteroides f...	795	0.0	G
ref|YP_212118.1| hypothetical protein BF2494 [Bacteroides fra...	795	0.0	G
ref|ZP_05281526.1| TPR repeat-containing protein [Bacteroides...	775	0.0	
ref|ZP_05417548.2| tetratricopeptide repeat family protein [B...	624	9e-177	
ref|ZP_01958966.1| hypothetical protein BACCAC_00554 [Bactero...	621	5e-176	
ref|ZP_05756644.1| TPR repeat-containing protein [Bacteroides...	620	1e-175	
ref|ZP_02063765.1| hypothetical protein BACOVA_00723 [Bactero...	620	1e-175	
ref|ZP_04546993.1| TPR repeat-containing protein [Bacteroides...	617	9e-175	
ref|ZP_04845680.1| TPR repeat-containing protein [Bacteroides...	598	3e-169	
ref|NP_809813.1| TPR repeat-containing protein [Bacteroides t...	598	3e-169	G
ref|ZP_03459917.1| hypothetical protein BACEGG_02718 [Bactero...	596	2e-168	
ref|ZP_02435236.1| hypothetical protein BACSTE_01478 [Bactero...	592	2e-167	
ref|ZP_06202893.1| conserved hypothetical protein [Bacteroide...	586	1e-165	
ref|ZP_02070574.1| hypothetical protein BACUNI_01995 [Bactero...	586	2e-165	
ref|ZP_03680531.1| hypothetical protein BACCELL_04904 [Bacter...	585	4e-165	
ref|ZP_03014415.1| hypothetical protein BACINT_01988 [Bactero...	567	9e-160	
ref|ZP_03010137.1| hypothetical protein BACCOP_02007 [Bactero...	256	3e-66	
ref|ZP_03644242.1| hypothetical protein BACCOPRO_02619 [Bacte...	247	2e-63	
ref|YP_001300548.1| TPR repeat-containing protein [Bacteroide...	247	2e-63	G
ref|ZP_03301317.1| hypothetical protein BACDOR_02699 [Bactero...	246	2e-63	
ref|ZP_05254725.1| TPR domain-containing protein [Bacteroides...	246	5e-63	
ref|ZP_06074897.1| TPR domain-containing protein [Bacteroides...	240	2e-61	
ref|YP_001302476.1| TPR domain-containing protein [Parabacter...	238	7e-61	G
ref|ZP_05545502.1| TPR domain-containing protein [Parabactero...	238	7e-61	
ref|ZP_03207140.1| hypothetical protein BACPLE_00760 [Bactero...	238	1e-60	
ref|ZP_06253438.1| putative TPR domain protein [Prevotella co...	230	2e-58	
ref|ZP_06005642.2| conserved hypothetical protein [Prevotella...	211	2e-52	
ref|ZP_06255209.1| putative TPR domain protein [Prevotella or...	206	5e-51	
ref|ZP_04833032.1| putative TPR domain protein [Prevotella me...	205	9e-51	
ref|ZP_06290121.1| conserved hypothetical protein [Prevotella...	198	7e-49	
ref|ZP_05857835.1| putative TPR domain protein [Prevotella ve...	197	2e-48	
ref|ZP_04055111.1| TPR domain protein [Porphyromonas uenonis ...	195	7e-48	
ref|ZP_02031759.1| hypothetical protein PARMER_01764 [Parabac...	190	2e-46	
ref|YP_001928992.1| TPR domain protein [Porphyromonas gingiva...	186	5e-45	G
ref|NP_905547.1| TPR domain-containing protein [Porphyromonas...	184	2e-44	G
ref|ZP_06287662.1| conserved hypothetical protein [Prevotella...	182	7e-44	
ref|ZP_06269395.1| conserved hypothetical protein [Prevotella...	181	2e-43	
ref|ZP_03475728.1| hypothetical protein PRABACTJOHN_01391 [Pa...	171	2e-40	
ref|ZP_05918007.1| conserved hypothetical protein [Prevotella...	168	8e-40	
ref|ZP_05369116.1| TPR domain protein [Propionibacterium sp. ...	147	2e-33	
ref|ZP_04390828.1| TPR domain protein [Porphyromonas endodont...	147	2e-33	
ref|ZP_05734739.1| putative tetratricopeptide repeat-containi...	137	3e-30	
ref|ZP_06255207.1| putative TPR domain protein [Prevotella or...	132	8e-29	
ref|ZP_06288289.1| tetratricopeptide repeat protein [Prevotel...	122	9e-26	
ref|ZP_05918006.1| conserved hypothetical protein [Prevotella...	114	2e-23	
ref|ZP_05734740.1| putative TPR domain protein [Prevotella ta...	108	8e-22	
ref|ZP_05857833.1| putative TPR domain protein [Prevotella ve...	108	9e-22	
ref|ZP_06287663.1| tetratricopeptide repeat protein [Prevotel...	106	4e-21	
ref|ZP_04833030.1| putative TPR domain protein [Prevotella me...	104	2e-20	
ref|ZP_06269393.1| tetratricopeptide repeat protein [Prevotel...	100	2e-19	
ref|ZP_06255208.1| hypothetical protein HMPREF0971_01238 [Pre...	100	4e-19	
ref|YP_100914.1| hypothetical protein BF3637 [Bacteroides fra...	95.5	9e-18	G
ref|ZP_04843481.1| conserved hypothetical protein [Bacteroide...	95.5	1e-17	
ref|ZP_03676726.1| hypothetical protein BACCELL_01053 [Bacter...	94.0	2e-17	
ref|ZP_05283200.1| hypothetical protein Bfra3_18177 [Bacteroi...	92.4	8e-17	
ref|ZP_06005641.2| conserved hypothetical protein [Prevotella...	89.7	5e-16	
ref|ZP_03209390.1| hypothetical protein BACPLE_03064 [Bactero...	89.4	7e-16	
ref|ZP_04833031.1| TPR domain protein [Prevotella melaninogen...	87.4	3e-15	
ref|ZP_06251203.1| putative TPR domain protein [Prevotella co...	84.7	2e-14	
ref|ZP_05857834.1| TPR domain protein [Prevotella veroralis F...	83.6	4e-14	
ref|ZP_01718101.1| hypothetical protein ALPR1_12965 [Algoriph...	67.0	4e-09	
ref|ZP_02424158.1| hypothetical protein ALIPUT_00273 [Alistip...	63.2	6e-08	
ref|ZP_06269394.1| tetratricopeptide repeat protein [Prevotel...	62.8	6e-08	
ref|YP_003290272.1| Tetratricopeptide TPR_2 repeat protein [R...	62.4	1e-07	G
ref|YP_861545.1| TPR repeat-containing protein [Gramella fors...	59.7	6e-07	G
ref|ZP_02177808.1| hypothetical protein HG1285_15791 [Hydroge...	59.3	8e-07	
pdb|1NA0|A Chain A, Design Of Stable Alpha-Helical Arrays Fro...	58.2	2e-06	S
ref|YP_001933073.1| hypothetical protein TPASS_0067 [Treponem...	57.8	2e-06	G
ref|NP_218507.1| hypothetical protein TP0067 [Treponema palli...	57.8	2e-06	G
ref|ZP_05623358.1| TPR domain protein [Treponema vincentii AT...	56.2	6e-06	
ref|NP_972674.1| TPR domain-containing protein [Treponema den...	55.8	9e-06	G
ref|YP_003125075.1| hypothetical protein Cpin_5445 [Chitinoph...	55.8	9e-06	G
ref|YP_001156602.1| TPR repeat-containing protein [Polynucleo...	55.1	1e-05	G
ref|YP_001952243.1| Tetratricopeptide TPR_2 repeat protein [G...	54.7	2e-05	G
ref|ZP_01463944.1| adventurous gliding protein T [Stigmatella...	54.7	2e-05	
ref|YP_720410.1| TPR repeat-containing protein [Trichodesmium...	54.7	2e-05	G
pdb|2FO7|A Chain A, Crystal Structure Of An 8 Repeat Consensu...	54.7	2e-05	S
ref|YP_002493579.1| TPR repeat-containing protein [Anaeromyxo...	54.3	2e-05	G
ref|XP_001026739.1| TPR Domain containing protein [Tetrahymen...	53.9	3e-05	G
ref|YP_466187.1| TPR repeat-containing protein [Anaeromyxobac...	53.9	3e-05	G
ref|YP_001996298.1| Tetratricopeptide TPR_2 repeat protein [C...	53.9	4e-05	G
ref|ZP_01048923.1| TPR repeat protein [Dokdonia donghaensis M...	53.1	5e-05	

ref|ZP_01253714.1|	TPR repeat [Psychroflexus torquis ATCC 700...	53.1	5e-05	
ref|YP_591514.1|	TPR repeat-containing protein [Candidatus Ko...	53.1	6e-05	G
ref|XP_001011832.1|	DNA polymerase family B containing protei...	52.8	8e-05	G
ref|YP_002730293.1|	tetratricopeptide repeat domain protein [...	52.8	8e-05	G
ref|ZP_01118022.1|	Tetratricopeptide repeat family protein [P...	52.8	8e-05	
ref|YP_002135421.1|	Tetratricopeptide TPR_2 repeat protein [A...	52.4	1e-04	G
ref|YP_002729512.1|	tetratricopeptide repeat domain protein [...	52.0	1e-04	G
ref|YP_003092431.1|	TPR repeat-containing protein [Pedobacter...	51.2	2e-04	G
ref|ZP_01251826.1|	hypothetical protein P700755_18349 [Psychr...	51.2	2e-04	
ref|YP_844212.1|	TPR repeat-containing protein [Syntrophobact...	50.8	3e-04	G
ref|ZP_01545479.1|	TPR repeat [Stappia aggregata IAM 12614] >...	50.8	3e-04	
ref|XP_001029949.2|	SLEI family protein [Tetrahymena thermoph...	50.8	3e-04	U G
ref|YP_002249256.1|	TPR domain protein, putative [Thermodesul...	50.4	3e-04	G
ref|YP_002603662.1|	TPR repeat protein [Desulfobacterium auto...	50.1	4e-04	G
ref|ZP_02736216.1|	hypothetical protein GobsU_30680 [Gemmata ...	50.1	4e-04	
ref|YP_306094.1|	TPR repeat-containing protein [Methanosarcin...	49.7	5e-04	G
ref|ZP_04780592.1|	tetratricopeptide repeat family protein [S...	49.7	6e-04	
ref|XP_314265.3|	AGAP003363-PA [Anopheles gambiae str. PEST] ...	49.7	6e-04	U G

Alignments Select All Get selected sequences Distance tree of results Multiple alignment NEW

```
>ref|YP_099695.1| G TPR repeat-containing protein [Bacteroides fragilis YCH46]
 ref|ZP_04843106.1|  TPR repeat-containing protein [Bacteroides sp. 3_2_5]
 ref|ZP_06092595.1|  TPR repeat-containing protein [Bacteroides sp. 2_1_16]
 dbj|BAD49161.1| G putative tetratricopeptide repeat family protein [Bacteroides
fragilis YCH46]
 gb|EES86337.1|  TPR repeat-containing protein [Bacteroides sp. 3_2_5]
 gb|EEZ27981.1|  TPR repeat-containing protein [Bacteroides sp. 2_1_16]
Length=403

 GENE ID: 3083889 BF2412 | TPR repeat-containing protein
[Bacteroides fragilis YCH46] (10 or fewer PubMed links)

 Score =  795 bits (2054),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 396/396 (100%), Positives = 396/396 (100%), Gaps = 0/396 (0%)

Query  1    MVLLMAVSFAFAQEKNVKEAKSIAGEVKPDFAKAEQLINEALTNPETKDNAATWDVAGYI  60
            MVLLMAVSFAFAQEKNVKEAKSIAGEVKPDFAKAEQLINEALTNPETKDNAATWDVAGYI
Sbjct  8    MVLLMAVSFAFAQEKNVKEAKSIAGEVKPDFAKAEQLINEALTNPETKDNAATWDVAGYI  67

Query  61   QKRINEKEMENAYLRKPYDTLKVYNSVLNMYNYYVKCDELAQIPNEKGKIKNKYRSANSK  120
            QKRINEKEMENAYLRKPYDTLKVYNSVLNMYNYYVKCDELAQIPNEKGKIKNKYRSANSK
Sbjct  68   QKRINEKEMENAYLRKPYDTLKVYNSVLNMYNYYVKCDELAQIPNEKGKIKNKYRSANSK  127

Query  121  TILAERPNLINGGIQYFNLNKNEDALKYFAAYVDAATLPMMEKENLLEKDTILPQVAYYA  180
            TILAERPNLINGGIQYFNLNKNEDALKYFAAYVDAATLPMMEKENLLEKDTILPQVAYYA
Sbjct  128  TILAERPNLINGGIQYFNLNKNEDALKYFAAYVDAATLPMMEKENLLEKDTILPQVAYYA  187

Query  181  TLAADRVGDKDAVMKYAQYALKDKENGQFAMQLLTDAYKAKGDTAKWVEKLQEGIVKFPE  240
            TLAADRVGDKDAVMKYAQYALKDKENGQFAMQLLTDAYKAKGDTAKWVEKLQEGIVKFPE
Sbjct  188  TLAADRVGDKDAVMKYAQYALKDKENGQFAMQLLTDAYKAKGDTAKWVEKLQEGIVKFPE  247

Query  241  NQYFFANLVDYYSSSNQNDKAMQFADDMLAKDPNNKLYLYVKAYLYHNMKDYEKAIEFYK  300
            NQYFFANLVDYYSSSNQNDKAMQFADDMLAKDPNNKLYLYVKAYLYHNMKDYEKAIEFYK
Sbjct  248  NQYFFANLVDYYSSSNQNDKAMQFADDMLAKDPNNKLYLYVKAYLYHNMKDYEKAIEFYK  307

Query  301  KTLDIDPAYAEACSNLGLVYLLQAQEYADKAPADINDPNYATAQAEIKKFYEAAKPYYEK  360
            KTLDIDPAYAEACSNLGLVYLLQAQEYADKAPADINDPNYATAQAEIKKFYEAAKPYYEK
Sbjct  308  KTLDIDPAYAEACSNLGLVYLLQAQEYADKAPADINDPNYATAQAEIKKFYEAAKPYYEK  367

Query  361  ARELKPDQKDLWLQGLYRVYYNLNMGPEFEEIEKMM  396
            ARELKPDQKDLWLQGLYRVYYNLNMGPEFEEIEKMM
Sbjct  368  ARELKPDQKDLWLQGLYRVYYNLNMGPEFEEIEKMM  403

>ref|YP_212118.1| G hypothetical protein BF2494 [Bacteroides fragilis NCTC 9343]
 emb|CAH08194.1| G conserved hypothetical protein [Bacteroides fragilis NCTC 9343]
Length=396

 GENE ID: 3287481 BF2494 | hypothetical protein [Bacteroides fragilis NCTC 9343]
(10 or fewer PubMed links)

 Score =  795 bits (2054),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 396/396 (100%), Positives = 396/396 (100%), Gaps = 0/396 (0%)

Query  1    MVLLMAVSFAFAQEKNVKEAKSIAGEVKPDFAKAEQLINEALTNPETKDNAATWDVAGYI  60
            MVLLMAVSFAFAQEKNVKEAKSIAGEVKPDFAKAEQLINEALTNPETKDNAATWDVAGYI
Sbjct  1    MVLLMAVSFAFAQEKNVKEAKSIAGEVKPDFAKAEQLINEALTNPETKDNAATWDVAGYI  60

Query  61   QKRINEKEMENAYLRKPYDTLKVYNSVLNMYNYYVKCDELAQIPNEKGKIKNKYRSANSK  120
            QKRINEKEMENAYLRKPYDTLKVYNSVLNMYNYYVKCDELAQIPNEKGKIKNKYRSANSK
Sbjct  61   QKRINEKEMENAYLRKPYDTLKVYNSVLNMYNYYVKCDELAQIPNEKGKIKNKYRSANSK  120

Query  121  TILAERPNLINGGIQYFNLNKNEDALKYFAAYVDAATLPMMEKENLLEKDTILPQVAYYA  180
            TILAERPNLINGGIQYFNLNKNEDALKYFAAYVDAATLPMMEKENLLEKDTILPQVAYYA
Sbjct  121  TILAERPNLINGGIQYFNLNKNEDALKYFAAYVDAATLPMMEKENLLEKDTILPQVAYYA  180

Query  181  TLAADRVGDKDAVMKYAQYALKDKENGQFAMQLLTDAYKAKGDTAKWVEKLQEGIVKFPE  240
            TLAADRVGDKDAVMKYAQYALKDKENGQFAMQLLTDAYKAKGDTAKWVEKLQEGIVKFPE
Sbjct  181  TLAADRVGDKDAVMKYAQYALKDKENGQFAMQLLTDAYKAKGDTAKWVEKLQEGIVKFPE  240

Query  241  NQYFFANLVDYYSSSNQNDKAMQFADDMLAKDPNNKLYLYVKAYLYHNMKDYEKAIEFYK  300
            NQYFFANLVDYYSSSNQNDKAMQFADDMLAKDPNNKLYLYVKAYLYHNMKDYEKAIEFYK
Sbjct  241  NQYFFANLVDYYSSSNQNDKAMQFADDMLAKDPNNKLYLYVKAYLYHNMKDYEKAIEFYK  300

Query  301  KTLDIDPAYAEACSNLGLVYLLQAQEYADKAPADINDPNYATAQAEIKKFYEAAKPYYEK  360
            KTLDIDPAYAEACSNLGLVYLLQAQEYADKAPADINDPNYATAQAEIKKFYEAAKPYYEK
Sbjct  301  KTLDIDPAYAEACSNLGLVYLLQAQEYADKAPADINDPNYATAQAEIKKFYEAAKPYYEK  360

Query  361  ARELKPDQKDLWLQGLYRVYYNLNMGPEFEEIEKMM  396
            ARELKPDQKDLWLQGLYRVYYNLNMGPEFEEIEKMM
Sbjct  361  ARELKPDQKDLWLQGLYRVYYNLNMGPEFEEIEKMM  396
```

H Crystal screen compositions

1	50% (w/v) PEG 400	0.2 M Lithium Sulphate	0.1 M Sodium acetate pH 5.1	
2	20% (w/v) PEG 3000	0.1 M Sodium citrate pH 5.5		
3	20% (w/v) PEG 3350	0.2 M Diammonium hydrogen citrate pH 5.0		
4	30% (v/v) MPD	0.08 M Calcium chloride	0.1 M Sodium acetate pH 4.6	
5	20% (w/v) PEG 3350	0.2 M Magnesium formate pH 5.9		
6	20% (w/v) PEG 1000	0.2 M lithium sulphate	0.25 M Sodiumcitrate pH 4.2	0.25 M sodium dihydrogen phosphate pH 5.2
7	20% (w/v) PEG 8000	0.1M CHES pH 9.5		
8	20% (w/v) PEG 3350	0.2 M Ammonium formate pH6.6		
9	20% (w/v) PEG 3350	0.2 M Ammonium chloride pH 6.3		
10	20% (w/v) PEG 3350	0.2 M Potassium formate pH 7.3		
11	50% (v/v) MPD	0.2 M Ammonium dihydrogen phosphate	0.1M Tris pH 8.5	
12	20% (w/v) PEG 3350	0.2 M Potassium nitrate pH 6.9		
13	0.8 M Ammonium sulphate	0.1M Citric acid pH 4.0		
14	20% (w/v) PEG 3350	0.2 M Sodium		

		thiocyanate pH 6.9		
15	20% (w/v) PEG 6000	0.1M Bicine pH 9.0		
16	10% (w/v) PEG 8000	8% (v/v) Ethylene glycol	0.1 M HEPES pH 7.5	
17	40% (v/v) MPD	5% (w/v) PEG 8000	0.1M Sodium cacodylate pH 7.0	
18	40% (v/v) Ethanol	5% (w/v) PEG 1000	0.25 M Sodium citrate pH 5.2	0.25 M Sodium dihydrogen phosphate pH 5.2
19	8% (w/v) PEG 4000	0.1M Sodium acetate pH 4.6		
20	10% (w/v) PEG 8000	0.2 M Magnesium chloride	0.1 M Tris pH 7.0	
21	20% (w/v) PEG 6000	0.1M Citric acid pH 5.0		
22	50% (w/v) PEG 200	0.2 M Magnesium chloride	0.1 M Sodium cacodylate pH 6.6	
23	1.6 M Sodium citrate pH 6.5			
24	20% (w/v) PEG 3350	0.2 M Potassium citrate pH 8.3		
25	30% (v/v) MPD	0.02 M Calcium chloride	0.1M Sodium acetate pH 4.6	
26	20% (w/v) PEG 8000	0.2 M Sodium chloride	0.25 M Sodium citrate pH 4.2	0.25 M Sodium dihydrogen phosphate pH 4.2
27	20% (w/v) PEG 6000	1.0 M Lithium chloride	0.1 M Citric acid pH 4.0	
28	20% (w/v) PEG 3350	0.2 M Ammonium nitrate pH 6.3		

29	10% (w/v) PEG 6000	0.1M HEPES pH 7.0		
30	0.8 M Ammonium dihydrogen phosphate	0.8 M Potassium dihydrogen phosphate	0.1M HEPES pH 7.5	
31	40% (w/v) PEG 300	0.25 M Sodium citrate pH 5.2	0.25 M Sodium dihydrogen phosphate pH 5.2	
32	10% (w/v) PEG 3000	0.2 M Zinc acetate	0.1 M Sodium acetate pH 4.5	
33	20% (v/v) Ethanol	0.1 M Tris pH 8.5		
34	25% (v/v)1-2-Propanediol	0.1M Sodium potassium phosphate pH 6.8	10% (v/v) glycerol	
35	10% (w/v) PEG 20000	2% (v/v) Dioxane	0.1M Bicine pH 9.0	
36	2.0 M Ammonium sulphate	0.1M Sodium acetate pH 4.6		
37	10% (w/v) PEG 1000	10% (w/v) PEG 8000		
38	24% (w/v) PEG 1000	20% (v/v) Glycerol		
39	30% (v/v) PEG 400	0.2 M Magnesium chloride	0.1M HEPES pH 7.5	
40	50% (w/v) PEG 200	0.2 M Sodium chloride	0.1 M Sodium potassium phosphate pH 7.2	
41	30% (w/v) PEG 8000	0.2 M Lithium sulphate	0.1M Sodium acetate pH 4.5	
42	70% (v/v) MPD	0.2 M Magnesium chloride	0.1 M HEPES pH 7.5	
43	20% (w/v) PEG 8000	0.1M Tris pH 8.5		

44	40% (v/v) PEG 400	0.2 M Lithium sulphate	0.1 M Tris pH 8.4	
45	40%(v/v) MPD	0.1M Tris pH 8.0		
46	25.5% (w/v) PEG 4000	0.17 M Ammonium sulphate	15% (v/v) Glycerol	
47	40% (w/v) PEG 300	0.2 M Calcium acetate	0.1M Sodium cacodylate pH 7.0	
48	14% (v/v) Isopropanol	0.14 M Calcium chloride	0.07M Sodium acetate pH 4.6	30% (v/v) glycerol
49	16% (w/v)PEG 8000	0.04 M Potassium dihydrogen phosphate	20% (v/v) Glycerol	
50	1.0 M Sodium citrate	0.1 M sodium cacodylate pH 6.5		
51	2.0 M Ammonium sulphate	0.2 M Sodium chloride	0.1 M Sodium cacodylate pH 6.5	
52	10% (v/v) Isopropanol	0.2 M Sodium chloride	0.1 M HEPES pH 7.5	
53	1.26 M Ammonium sulphate	0.2 M Lithium sulphate	0.1M Tris pH 8.5	
54	40% (v/v) MPD	0.1M CAPS pH 10.1		
55	20% (w/v)PEG 3000	0.2 M Zinc acetate	0.1M Imidazole pH 8.0	
56	10% (v/v) Isopropanol	0.2 M Zinc acetate	0.1M Sodium cacodylate pH 6.5	
57	1.0M Diammonium hydrogen phosphate	0.1 M Sodium acetate pH 4.5		
58	1.6 M Magnesium sulphate	0.1M MES pH 6.5		

59	10% (w/v)PEG 6000	0.1 M Bicine pH 9.0		
60	14.4% (w/v)PEG 8000	0.16 M Calcium acetate	0.08 M Sodium cacodylate pH 6.5	20% (v/v) glycerol
61	10% (w/v)PEG 8000	0.1 M Imidazol pH 8.0		
62	30% (v/v) Jeffamine	0.05 M Caesiumchloride	0.1 M MES pH 6.5	
63	3.2 M Ammonium sulphate	0.1 M Citric acid pH 5.0		
64	20% (v/v) MPD	0.1 M Tris pH 8.0		
65	20% (v/v) Jeffamine	0.1M HEPES pH 6.5		
66	50% (v/v) Ethylene glycol	0.2 M Magnesium chloride	0.1 M Tris pH 8.5	
67	10% (v/v) MPD	0.1 M Bicine pH 9.0		
68	0.2 M Ammonium sulphate	0.1 M Sodium acetate pH 4.6	30% (w/v) PEG MME 2000	
69	0.2 M Ammonium sulphate	0.1 M MES pH 6.5	30% (w/v) PEG MME 5000	
70	0.01 M Zinc sulphate	0.1 M MES pH 6.5	25% (w/v) PEG MME 550	
71	0.01 M Nickel chloride	0.1 M Tris pH 8.5	20% (w/v) PEG MME 2000	
72	0.1 M Sodium chloride	0.1 M Bicine pH 9.0	20% (w/v) PEG MME 550	
73	0.005 M Magnesium chloride	0.05 M HEPES pH 7.0	25% (w/v) PEG MME 550	

74	0.1 M Potassium chloride	0.015 M Magnesium chloride	0.05 M Tris pH 7.5	10% PEG MME 550
75	20% (v/v)1-4-Butandiol	0.1 M MES pH 6.0	0.2 M Lithium sulphate	
76	1 M Sodium potassium tartrate	0.1 M Imidazole pH 8.0	0.2 M Sodium chloride	
77	20% (v/v)1-4-Butandiol	0.1M Sodium acetate pH 4.5		
78	1M Sodium potassium tartrate	0.1 M CHES pH 9.5	0.2 M Lithium sulphate	
79	35% (v/v) Propanol	0.1 M Sodium cacodylate pH6.5		
80	35% (v/v) Propanol	0.1 M Tris pH 8.5		
81	3.5 M Sodium formate			
82	0.8M Succinic acid pH 7.0			
83	2.1 M Malic acid pH 7.0			
84	2.4 M Sodium malonate pH 7.0			
85	0.2 M Potassium chloride	0.05 M HEPES pH 7.5	35% (v/v) Pentaerythritol propoxylate	
86	0.005 M Ammonium sulphate	0.05 M Tris pH 6.5	30% (v/v) Pentaerythritol ethoxylate	
87	0.2 M Potassium	25% (w/v) PEG MME	0.1 M HEPES pH 7.5	

	bromide	2000		
88	0.2M Potassium bromide	8% (w/v)PEG 20000	8% (w/v) PEG MME 550	0.1 M Tris pH 8.5
89	1.0 M Potassium dihydrogen phosphate	0.1 M Sodium citrate pH 4.6		
90	0.5 M Potassium dihydrogen phosphate	0.1M HEPES pH 7.0		
91	20% (w/v)PEG 4000	0.005 M Cadmium chloride	0.1 M Tris pH 8.0	
92	20% (w/v)PEG 4000	0.005 M Nickel chloride	0.1 M MED pH 6.5	
93	0.8 M Sodium formate	10% (w/v) PEG 8000	10% PEG 1000	0.1 M imidazol pH 8.0
94	15% (w/v)PEG 4000	0.005 M Cadmium sulphate	0.1 M Sodium cacodylate pH 6.5	
95	20% (w/v) PEG 600	0.005 M Cobalt chloride	0.1 M HEPES pH 7.5	
96	2 M Ammonium sulphate	10% (v/v) Jeffamine	0.1 M Tris pH 8.0	

Newcastle Crystal screen formulation

PACT crystal screen formulations

PEG/Anion Screens

Well number	Salt	Buffer	PEG
1	0.2 M Sodium fluoride		20% (w/v) PEG 3350
2	0.2 M Sodium bromide		20% (w/v) PEG 3350
3	0.2 M Sodium iodide		20% (w/v) PEG 3350
4	0.2 M Potassium thiocyanate		20% (w/v) PEG 3350
5	0.2 M Sodium nitrate		20% (w/v) PEG 3350
6	0.2 M Sodium formate		20% (w/v) PEG 3350
7	0.2 M Sodium acetate		20% (w/v) PEG 3350
8	0.2 M Sodium sulphate		20% (w/v) PEG 3350
9	0.2 M Sodium potassium tartrate		20% (w/v) PEG 3350
10	0.2 M Sodium potassium phosphate		20% (w/v) PEG 3350
11	0.2 M Sodium citrate		20%(w/v) PEG 3350
12	0.2 M Sodium malonate		20% (w/v) PEG 3350
13	0.2 M Sodium fluoride	0.1 M Bis-tris propane pH 6.5	20% (w/v) PEG 3350
14	0.2 M Sodium bromide	0.1 M Bis-tris propane pH 6.5	20% (w/v) PEG 3350
15	0.2 M Sodium iodide	0.1 M Bis-tris propane pH 6.5	20% (w/v) PEG 3350
16	0.2 M Potassium thiocyanate	0.1 M Bis-tris propane pH 6.5	20% (w/v) PEG 3350
17	0.2 M Sodium nitrate	0.1 M Bis-tris propane pH 6.5	20% (w/v) PEG 3350
18	0.2 M Sodium formate	0.1 M Bis-tris propane pH 6.5	20% (w/v) PEG 3350
19	0.2 M Sodium acetate	0.1 M Bis-tris propane pH 6.5	20% (w/v) PEG 3350
20	0.2 M Sodium sulphate	0.1 M Bis-tris propane pH 6.5	20% (w/v) PEG 3350
21	0.2 M Sodium potassium tartrate	0.1 M Bis-tris propane pH 6.5	20% (w/v) PEG 3350
22	0.2 M Sodium potassium phosphate	0.1 M Bis-tris propane pH 6.5	20% (w/v) PEG 3350
23	0.2 M Sodium citrate	0.1 M Bis-tris propane pH 6.5	20% (w/v) PEG 3350
24	0.2 M Sodium malonate	0.1 M Bis-tris propane pH 6.5	20% (w/v) PEG 3350
25	0.2 M Sodium fluoride	0.1 M Bis-tris propane pH 7.5	20% (w/v) PEG 3350
26	0.2 M Sodium bromide	0.1 M Bis-tris propane pH 7.5	20% (w/v) PEG 3350
27	0.2 M Sodium iodide	0.1 M Bis-tris propane pH 7.5	20% (w/v) PEG 3350

28	0.2 M Potassium thiocyanate	0.1 M Bis-tris propane pH 7.5	20% (w/v) PEG 3350
29	0.2 M Sodium nitrate	0.1 M Bis-tris propane pH 7.5	20% (w/v) PEG 3350
30	0.2 M Sodium formate	0.1 M Bis-tris propane pH 7.5	20% (w/v) PEG 3350
31	0.2 M Sodium acetate	0.1 M Bis-tris propane pH 7.5	20% (w/v) PEG 3350
32	0.2 M Sodium sulphate	0.1 M Bis-tris propane pH 7.5	20% (w/v) PEG 3350
33	0.2 M Sodium potassium tartrate	0.1 M Bis-tris propane pH 7.5	20% (w/v) PEG 3350
34	0.2 M Sodium potassium phosphate	0.1 M Bis-tris propane pH 7.5	20% (w/v) PEG 3350
35	0.2 M Sodium citrate	0.1 M Bis-tris propane pH 7.5	20% (w/v) PEG 3350
36	0.2 M Sodium malonate	0.1 M Bis-tris propane pH 7.5	20% (w/v) PEG 3350
37	0.2 M Sodium fluoride	0.1 M Bis-tris propane pH 8.5	20% (w/v) PEG 3350
38	0.2 M Sodium bromide	0.1 M Bis-tris propane pH 8.5	20% (w/v) PEG 3350
39	0.2 M Sodium iodide	0.1 M Bis-tris propane pH 8.5	20% (w/v) PEG 3350
40	0.2 M Potassium thiocyanate	0.1 M Bis-tris propane pH 8.5	20% (w/v) PEG 3350
41	0.2 M Sodium nitrate	0.1 M Bis-tris propane pH 8.5	20% (w/v) PEG 3350
42	0.2 M Sodium formate	0.1 M Bis-tris propane pH 8.5	20% (w/v) PEG 3350
43	0.2 M Sodium acetate	0.1 M Bis-tris propane pH 8.5	20% (w/v) PEG 3350
44	0.2 M Sodium sulphate	0.1 M Bis-tris propane pH 8.5	20% (w/v) PEG 3350
45	0.2 M Sodium potassium tartrate	0.1 M Bis-tris propane pH 8.5	20% (w/v) PEG 3350
46	0.2 M Sodium potassium phosphate	0.1 M Bis-tris propane pH 8.5	20% (w/v) PEG 3350
47	0.2 M Sodium citrate	0.1 M Bis-tris propane pH 8.5	20% (w/v) PEG 3350
48	0.2 M Sodium malonate	0.1 M Bis-tris propane pH 8.5	20% (w/v) PEG 3350

PEG/Cation Screens

Well number		Buffer	PEG
1	0.2 M Sodium chloride	Acetate pH 5	20% (w/v) PEG 6000
2	0.2 M Ammonium chloride	Acetate pH 5	20% (w/v) PEG 6000
3	0.2 M Lithium chloride	Acetate pH 5	20% (w/v) PEG 6000
4	0.2 M Magnesium chloride	Acetate pH 5	20% (w/v) PEG 6000
5	0.2 M Calcium chloride	Acetate pH 5	20% (w/v) PEG 6000
6	0.01 M Zinc chloride	Acetate pH 5	20% (w/v) PEG 6000
7	0.2 M Sodium chloride	MES pH 6	20% (w/v) PEG 6000
8	0.2 M Ammonium chloride	MES pH 6	20% (w/v) PEG 6000
9	0.2 M Lithium chloride	MES pH 6	20% (w/v) PEG 6000
10	0.2 M Magnesium chloride	MES pH 6	20% (w/v) PEG 6000
11	0.2 M Calcium chloride	MES pH 6	20%(w/v) PEG 6000
12	0.01 M Zinc chloride	MES pH 6	20% (w/v) PEG 6000
13	0.2 M Sodium chloride	HEPES pH 7	20% (w/v) PEG 6000
14	0.2 M Ammonium chloride	HEPES pH 7	20% (w/v) PEG 6000
15	0.2 M Lithium chloride	HEPES pH 7	20% (w/v) PEG 6000
16	0.2 M Magnesium chloride	HEPES pH 7	20% (w/v) PEG 6000
17	0.2 M Calcium chloride	HEPES pH 7	20% (w/v) PEG 6000
18	0.01 M Zinc chloride	HEPES pH 7	20% (w/v) PEG 6000
19	0.2 M Sodium chloride	Tris pH 8	20% (w/v) PEG 6000
20	0.2 M Ammonium chloride	Tris pH 8	20% (w/v) PEG 6000
21	0.2 M Lithium chloride	Tris pH 8	20% (w/v) PEG 6000
22	0.2 M Magnesium chloride	Tris pH 8	20% (w/v) PEG 6000
23	0.2 M Calcium chloride	Tris pH 8	20% (w/v) PEG 6000
24	0.01 M Zinc chloride	Tris pH 8	20% (w/v) PEG 6000

PEG/pH Screen compositions

Well number	Buffer system	pH of the buffer system	PEG
1	1 M Sodium malonate, 1 M imidazole, 1 M Boric acid	4	25% (w/v) PEG 1500
2	1 M Sodium malonate, 1 M imidazole, 1 M Boric acid	5	25% (w/v) PEG 1500
3	1 M Sodium malonate, 1 M imidazole, 1 M Boric acid	6	25% (w/v) PEG 1500
4	1 M Sodium malonate, 1 M imidazole, 1 M Boric acid	7	25% (w/v) PEG 1500
5	1 M Sodium malonate, 1 M imidazole, 1 M Boric acid	8	25% (w/v) PEG 1500
6	1 M Sodium malonate, 1 M imidazole, 1 M Boric acid	9	25% (w/v) PEG 1500
7	1 M Sodium propionate, 1 M Sodium cacodylate trihydrate, 1 M Bis tris propane	4	25% (w/v) PEG 1500
8	1 M Sodium propionate, 1 M Sodium cacodylate trihydrate, 1 M Bis tris propane	5	25% (w/v) PEG 1500
9	1 M Sodium propionate, 1 M Sodium cacodylate trihydrate, 1 M Bis tris propane	6	25% (w/v) PEG 1500
10	1 M Sodium propionate, 1 M Sodium cacodylate trihydrate, 1 M Bis tris propane	7	25% (w/v) PEG 1500
11	1 M Sodium propionate, 1 M Sodium cacodylate trihydrate, 1 M Bis tris propane	8	25% (w/v) PEG 1500
12	1 M Sodium propionate, 1 M Sodium cacodylate trihydrate, 1 M Bis tris propane	9	25% (w/v) PEG 1500
13	1 M DL-Malic acid, 1 M MES, 1 M Tris	4	25% (w/v) PEG 1500
14	1 M DL-Malic acid, 1 M MES, 1 M Tris	5	25% (w/v) PEG 1500
15	1 M DL-Malic acid, 1 M MES, 1 M Tris	6	25% (w/v) PEG 1500

16	1 M DL-Malic acid, 1 M MES, 1 M Tris	7	25% (w/v) PEG 1500
17	1 M DL-Malic acid, 1 M MES, 1 M Tris	8	25% (w/v) PEG 1500
18	1 M DL-Malic acid, 1 M MES, 1 M Tris	9	25% (w/v) PEG 1500

Clear Strategy Screen 1 Conditions (CSS 1)

1 0.3 M Sodium acetate, 25% (w/v) PEG 2000 MME	2 0.2 M Lithium sulphate, 25%(w/v) PEG 2000 MME	3 0.2 M Magnesium chloride, 25% (w/v) PEG 2000 MME	4 0.2 M Potassium bromide, 25% (w/v) PEG 2000 MME	5 0.2 M Potassium thiocyanate, 25% (w/v) PEG 2000 MME	6 0.8 M Sodium formate, 25% (w/v) PEG 2000 MME
7 0.3 M Sodium acetate, 15% (w/v) PEG 4000	8 0.2 M Lithium sulphate, 15% (w/v) PEG 4000	9 0.2 M Magnesium chloride, 15% (w/v) PEG 4000	10 0.2 M Potassium bromide, 15% (w/v) PEG 4000	11 0.2 M Potassium thiocyanate, 15% (w/v) PEG 4000	12 0.8 M Sodium formate, 15% (w/v) PEG 4000
13 0.3 M Sodium acetate, 10% (w/v) PEG 8000, 10% (w/v) PEG 1000	14 0.2 M Lithium sulphate, 10% (w/v) PEG 8000, 10% (w/v) PEG 1000	15 0.2 M Magnesium chloride, 10% (w/v) PEG 8000, 10% (w/v) PEG 1000	16 0.2 M Potassium bromide, 10% (w/v) PEG 8000, 10% (w/v) PEG 1000	17 0.2 M Potassium thiocyanate, 10% (w/v) PEG 8000, 10% (w/v) PEG 1000	18 0.8 M Sodium formate, 10% (w/v) PEG 8000, 10% (w/v) PEG 1000
19 0.3 M Sodium acetate 8% (w/v) PEG 20000, 8% (w/v) PEG550 MME	20 0.2 M Lithium sulphate, 8% (w/v) PEG 20000, 8% (w/v) PEG550 MME	21 0.2 M Magnesium chloride, 8% (w/v) PEG 20000, 8% (w/v) PEG550 MME	22 0.2 M Potassium bromide, 8% (w/v) PEG 20000, 8% (w/v) PEG550 MME	23 0.2 M Potassium thiocyanate, 8% (w/v) PEG 20000, 8% (w/v) PEG550 MME	24 0.8 M Sodium formate, 8% (w/v) PEG 20000, 8% (w/v) PEG550 MME

Clear Strategy Screen 2 conditions (MDCS2)

[A1] Tube : 1 1.5 M Ammonium sulphate	[A2] Tube : 2 0.8 M Lithium sulphate	[A3] Tube : 3 2 M Sodium formate	[A4] Tube : 4 0.5 M Potassium phosphate	[A5] Tube : 5 25% (w/v) PEG 2000 MME, 0.2 M Calcium acetate	[A6] Tube : 6 15% (w/v) PEG 4000, 0.2 M Calcium acetate
[B1] Tube : 7 2.7 M Ammonium sulphate	[B2] Tube : 8 1.8 M Lithium sulphate	[B3] Tube : 9 4 M Sodium formate	[B4] Tube : 10 1.0 M Potassium dihydrogen phosphate	[B5] Tube : 11 10% (w/v) PEG 8000, 10% (w/v) PEG 1000, 0.2 M Calcium acetate	[B6] Tube : 12 8% (w/v) PEG 20000, 8% (w/v) PEG 550 MME, 0.2 M Calcium acetate
[C1] Tube : 13 40%v/v Methanepentanediol	[C2] Tube : 14 40%v/v Butanediol	[C3] Tube : 15 20% (w/v) PEG 4000, 0.005 M Cadmium chloride	[C4] Tube : 16 20% (w/v) PEG 550 MME, 0.15 M Potassium thiocyanate	[C5] Tube : 17 20% (w/v) PEG 600, 0.15 M Potassium thiocyanate	[C6] Tube : 18 20% (w/v) PEG 15000, 0.15 M Potassium thiocyanate
[D1] Tube : 19 35%v/v Isopropanol	[D2] Tube : 20 30% (v/v) Jeffamine 600M	[D3] Tube : 21 20% (w/v) PEG 4000, 0.005 M Nickel chloride	[D4] Tube : 22 18% (w/v) PEG 3350, 0.15 M Potassium thiocyanate	[D5] Tube : 23 18% (w/v) PEG 5000 MME, 0.15 M Potassium thiocyanate	[D6] Tube : 24 15% (w/v) PEG 6000, 0.15 M Potassium thiocyanate

Hampton screen 1 formulations (HS 1)

Tube	SALT	BUFFER	Precipitant
1	0.02 M Calcium Chloride	0.1 M Sodium acetate pH 4.6	30% (w/v) 2-methyl-2,4-pentanediol
2	None	None	0.4 M Potassium sodium tartrate tetrahydrate
3	None	None	0.4 M Ammonium dihydrogen phosphate
4	None	0.1 M Tris-HCl pH 8.5	2.0 M Ammonium Sulphate
5	0.2 M tri-sodium citrate	0.1 M sodium HEPES pH 7.5	30% (w/v) 2-methyl-2,4-pentanediol
6	0.2 M Magnesium chloride	0.1 M Tris-HCl pH 8.5	30% (w/v) PEG 4000
7	None	0.1 M sodium cacodylate pH 6.5	1.4 M Sodium acetate trihydrate
8	0.2 M tri-sodium citrate	0.1 M sodium cacodylate pH 6.5	30% (v/v) 2-propanol
9	0.2 M Ammonium acetate	0.1 M sodium citrate pH 5.6	30% (w/v) PEG 4000
10	0.2 M Ammonium acetate	0.1 M sodium acetate pH 4.6	30% (w/v) PEG 4000
11	None	0.1 M sodium citrate pH 5.6	1.0 M Ammonium dihydrogen phosphate
12	0.2 M Magnesium chloride	0.1 M sodium HEPES pH 7.5	30% (v/v) 2-propanol
13	0.2 M tri-sodium citrate	0.1 M Tris-HCl pH 8.5	30% (v/v) PEG 400
14	0.2 M Calcium Chloride	0.1 M sodium HEPES pH 7.5	28% (v/v) PEG 400
15	0.2 M Ammonium acetate	0.1 M sodium cacodylate pH 6.5	30% (w/v) PEG 8000
16	None	0.1 M sodium HEPES pH 7.5	1.5 M Lithium sulphate monohydrate
17	0.2 M Lithium sulphate	0.1 M Tris-HCl pH 8.5	30% (w/v) PEG 4000
18	0.2 M Magnesium acetate	0.1 M sodium cacodylate pH 6.5	20% (w/v) PEG 8000
19	0.2 M Ammonium acetate	0.1 M Tris-HCl pH 8.5	30% (v/v) 2-propanol
20	0.2 M Ammonium sulphate	0.1 M sodium acetate pH 4.6	25% (w/v) PEG 4000
21	0.2 M Magnesium acetate	0.1 M sodium cacodylate pH 6.5	30% (v/v) 2-methyl-2,4-pentanediol
22	0.2 M Sodium acetate	0.1 M Tris-HCl pH 8.5	30% (w/v) PEG 4000
23	0.2 M Magnesium chloride	0.1 M sodium HEPES pH 7.5	30% (v/v) PEG 400
24	0.2 M Calcium Chloride	0.1 M sodium acetate pH 4.6	20% (v/v) 2-propanol
25	None	0.1 M Imidazole pH 6.5	1.0 M Sodium acetate trihydrate
26	0.2 M Ammonium acetate	0.1 M sodium citrate pH 5.6	30% (v/v) 2-methyl-2,4-pentanediol
27	0.2 M tri-sodium citrate	0.1 M sodium HEPES pH 7.5	20% (v/v) 2-propanol
28	0.2 M Sodium acetate	0.1 M sodium Cacodylate pH 6.5	30% (w/v) PEG 8000
29	None	0.1 M sodium HEPES pH 7.5	0.8 M Potassium sodium tartrate
30	0.2 M Ammonium acetate	None	30% (w/v) PEG 8000
31	0.2 M Ammonium acetate	None	30% (w/v) PEG 4000
32	None	None	2.0 M Ammonium Sulphate
33	None	None	4.0 M Sodium formate
34	None	0.1 M sodium Acetate pH 4.6	2.0 M Sodium formate

35	None	0.1 M sodium HEPES pH 7.5	1.6 M Potassium sodium tartrate
36	None	0.1 M Tris-HCl pH 8.5	8% (w/v) PEG 8000
37	None	0.1 M sodium acetate pH 4.6	8% (w/v) PEG 4000
38	None	0.1 M sodium HEPES pH 7.5	1.4 M Sodium citrate dihydrate
39	None	0.1 M sodium HEPES pH 7.5	2% (v/v) PEG 400 & 2.0 M Ammonium Sulphate
40	None	0.1 M sodium citrate pH 5.6	20% (v/v) 2-propanol & 20% (w/v) PEG 4000
41	None	0.1 M sodium HEPES pH 7.5	10% (v/v) 2-propanol & 20% (w/v) PEG 4000
42	0.05 M potassium phosphate monobasic	None	20% (w/v) PEG 8000
43	None	None	30% (w/v) PEG 1500
44	None	None	0.2 M Magnesium formate
45	0.2 M Zinc acetate	0.1 M sodium cacodylate pH 6.5	18% (w/v) PEG 8000
46	0.2 M Calcium acetate	0.1 M sodium cacodylate pH 6.5	18% (w/v) PEG 8000
47	None	0.1 M sodium acetate pH 4.6	2.0 M Ammonium Sulphate
48	None	0.1 M Tris-HCl pH 8.5	2.0 M Ammonium dihydrogen phosphate

Hampton Crystal Screen II formulations (HCS2)

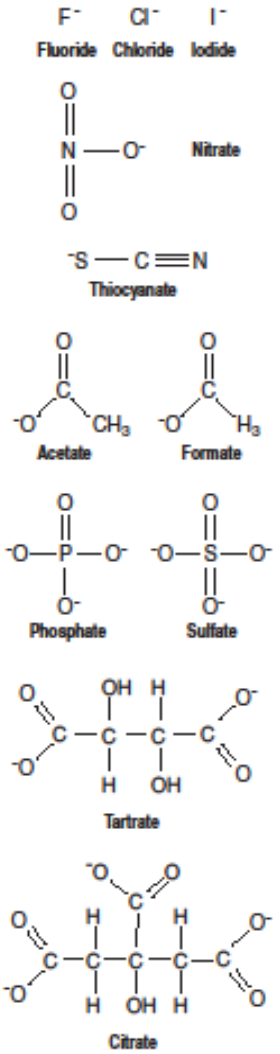
[A1] Tube : 1 2.0 M sodium chloride, 10% PEG 6000	[A2] Tube : 2 0.5 M sodium chloride, 0.01 M hexadecyltrimethylammonium bromide, 0.01 M magnesium chloride hexahydrate	[A3] Tube : 3 25% (v/v) ethylene glycol	[A4] Tube : 4 35% (v/v) dioxane	[A5] Tube : 5 2.0 M ammonium sulphate, 5% (v/v) iso-propanol
[B1] Tube : 7 10% (w/v) PEG 1000, 10% (w/v) PEG 8000	[B2] Tube : 8 1.5 M sodium chloride, 10% (v/v) ethanol	[B3] Tube : 9 0.1 M sodium acetate trihydrate pH 4.6, 2 M sodium chloride	[B4] Tube : 10 0.2 M sodium chloride, 0.1 M sodium acetate trihydrate pH 4.6, 30% (v/v) MPD	[B5] Tube : 11 0.01 M cobalt chloride hexahydrate, 0.1 M sodium acetate trihydrate pH 4.6, 1 M 1,6 hexanediol
[C1] Tube : 13 0.2 M ammonium sulphate, 0.1 M sodium acetate trihydrate pH 4.6, 30% (w/v) PEG monomethyl ether 2000	[C2] Tube : 14 0.2 M potassium sodium tartrate tetrahydrate, 0.1 M tri-sodium citrate dihydrate pH 5.6, 2 M ammonium sulfate	[C3] Tube : 15 0.5 M ammonium sulphate, 0.1 M tri-sodium citrate dihydrate pH 5.6, 1 M lithium sulfate monohydrate	[C4] Tube : 16 0.5 M sodium chloride, 0.1 M tri-sodium citrate dihydrate pH 5.6, 2% (v/v) ethylene imine polymer	[C5] Tube : 17 0.1 M tri-sodium citrate dihydrate pH 5.6, 35% (v/v) tert-butanol
[D1] Tube : 19 0.1 M tri-sodium citrate dehydrate, pH 5.6 2.5 M 1,6 hexanediol	[D2] Tube : 20 0.1 M MES pH 6.5, 1.6 M magnesium sulfate heptahydrate	[D3] Tube : 21 0.1 M sodium dihydrogen phosphate, 0.1 M MES pH 6.5, 2 M sodium chloride, 0.1 M potassium dihydrogen phosphate	[D4] Tube : 22 0.1 M MES pH 6.5, 12% (w/v) PEG 20,000	[D5] Tube : 23 1.6 M ammonium sulphate, 0.1 M MES pH 6.5 10% (v/v) dioxane
[E1] Tube : 25 0.01 M cobaltous chloride hexahydrate, 0.1 M MES pH 6.5, 1.8 M ammonium sulfate	[E2] Tube : 26 0.2 M ammonium sulphate, 0.1 M MES pH 6.5, 30% (w/v) PEG monomethyl ether 5000	[E3] Tube : 27 0.01 M zinc sulfate heptahydrate, 0.1 M MES pH 6.5 25% (v/v) PEG monomethyl ether 550	[E4] Tube : 28 pH 6.5 1.6 M tri-sodium citrate dihydrate	[E5] Tube : 29 0.5 M ammonium sulphate, 0.1 M HEPES pH 7.5, 30% (v/v) MPD
[F1] Tube : 31 0.1 M HEPES pH 7.5, 20% (v/v) jeffamine M-600	[F2] Tube : 32 0.1 M sodium chloride, 0.1 M HEPES pH 7.5, 1.6 M ammonium sulfate	[F3] Tube : 33 0.1 M HEPES pH 7.5, 2 M ammonium formate	[F4] Tube : 34 0.05 M cadmium sulfate hydrate, 0.1 M HEPES	[F5] Tube : 35 0.1 M HEPES pH 7.5, 70% (v/v) MPD

			pH 7.5, 1 M sodium acetate trihydrate	
[G1] Tube : 37 0.1 M HEPES pH 7.5, 10% (w/v) PEG 8000	[G2] Tube : 38 0.1 M HEPES pH 7.5, 20% (w/v) PEG 10000, 8% (v/v) ethylene glycol	[G3] Tube : 39 0.2 M magnesium chloride hexahydrate, 0.1 M tris pH 8.5, 3.4 M 1,6 hexanediol	[G4] Tube : 40 0.1 M tris pH 8.5, 25% (v/v) tert-butanol	[G5] Tube : 41 0.01 M nickel (II) chloride hexahydrate, 0.1 M tris pH 8.5, 1 M lithium sulfate monohydrate
[H1] Tube : 43 0.2 M ammonium dihydrogen phosphate, 0.1 M tris pH 8.5, 50% (v/v) MPD	[H2] Tube : 44 0.1 M tris pH 8.5, 20% (v/v) ethanol	[H3] Tube : 45 0.01 M nickel (II) chloride hexahydrate, 0.1 M tris pH 8.5, 20% (w/v) PEG monomethyl ether 2000	[H4] Tube : 46 0.1 M sodium chloride, 0.1 M bicine pH 9.0, 20% (v/v) PEG monomethyl ether 550	[H5] Tube : 47 0.1 M bicine, pH 9.0, 2 M magnesium chloride hexahydrate

Recipe courtesy of Hampton Research

Reference website
http://xray.bmc.uu.se/markh/php/xtalscreens.php?func=lookup&screen_name=Expand+List

Tube #	Salt	Tube #	Polymer	Tube #	pH ◊
1.	0.2 M Sodium fluoride	1.	20% w/v Polyethylene glycol 3,350	1.	7.3
2.	0.2 M Potassium fluoride	2.	20% w/v Polyethylene glycol 3,350	2.	7.3
3.	0.2 M Ammonium fluoride	3.	20% w/v Polyethylene glycol 3,350	3.	6.2
4.	0.2 M Lithium chloride	4.	20% w/v Polyethylene glycol 3,350	4.	6.8
5.	0.2 M Magnesium chloride hexahydrate	5.	20% w/v Polyethylene glycol 3,350	5.	5.9
6.	0.2 M Sodium chloride	6.	20% w/v Polyethylene glycol 3,350	6.	6.9
7.	0.2 M Calcium chloride dihydrate	7.	20% w/v Polyethylene glycol 3,350	7.	5.1
8.	0.2 M Potassium chloride	8.	20% w/v Polyethylene glycol 3,350	8.	7.0
9.	0.2 M Ammonium chloride	9.	20% w/v Polyethylene glycol 3,350	9.	6.3
10.	0.2 M Sodium iodide	10.	20% w/v Polyethylene glycol 3,350	10.	7.0
11.	0.2 M Potassium iodide	11.	20% w/v Polyethylene glycol 3,350	11.	7.0
12.	0.2 M Ammonium iodide	12.	20% w/v Polyethylene glycol 3,350	12.	6.2
13.	0.2 M Sodium thiocyanate	13.	20% w/v Polyethylene glycol 3,350	13.	6.9
14.	0.2 M Potassium thiocyanate	14.	20% w/v Polyethylene glycol 3,350	14.	7.0
15.	0.2 M Lithium nitrate	15.	20% w/v Polyethylene glycol 3,350	15.	7.1
16.	0.2 M Magnesium nitrate hexahydrate	16.	20% w/v Polyethylene glycol 3,350	16.	5.9
17.	0.2 M Sodium nitrate	17.	20% w/v Polyethylene glycol 3,350	17.	6.8
18.	0.2 M Potassium nitrate	18.	20% w/v Polyethylene glycol 3,350	18.	6.8
19.	0.2 M Ammonium nitrate	19.	20% w/v Polyethylene glycol 3,350	19.	6.2
20.	0.2 M Magnesium formate dihydrate	20.	20% w/v Polyethylene glycol 3,350	20.	7.0
21.	0.2 M Sodium formate	21.	20% w/v Polyethylene glycol 3,350	21.	7.2
22.	0.2 M Potassium formate	22.	20% w/v Polyethylene glycol 3,350	22.	7.3
23.	0.2 M Ammonium formate	23.	20% w/v Polyethylene glycol 3,350	23.	6.6
24.	0.2 M Lithium acetate dihydrate	24.	20% w/v Polyethylene glycol 3,350	24.	7.9
25.	0.2 M Magnesium acetate tetrahydrate	25.	20% w/v Polyethylene glycol 3,350	25.	7.9
26.	0.2 M Zinc acetate dihydrate	26.	20% w/v Polyethylene glycol 3,350	26.	6.4
27.	0.2 M Sodium acetate trihydrate	27.	20% w/v Polyethylene glycol 3,350	27.	8.0
28.	0.2 M Calcium acetate hydrate	28.	20% w/v Polyethylene glycol 3,350	28.	7.5
29.	0.2 M Potassium acetate	29.	20% w/v Polyethylene glycol 3,350	29.	8.1
30.	0.2 M Ammonium acetate	30.	20% w/v Polyethylene glycol 3,350	30.	7.1
31.	0.2 M Lithium sulfate monohydrate	31.	20% w/v Polyethylene glycol 3,350	31.	6.0
32.	0.2 M Magnesium sulfate heptahydrate	32.	20% w/v Polyethylene glycol 3,350	32.	6.0
33.	0.2 M Sodium sulfate decahydrate	33.	20% w/v Polyethylene glycol 3,350	33.	6.7
34.	0.2 M Potassium sulfate	34.	20% w/v Polyethylene glycol 3,350	34.	6.8
35.	0.2 M Ammonium sulfate	35.	20% w/v Polyethylene glycol 3,350	35.	6.0
36.	0.2 M Sodium tartrate dibasic dihydrate	36.	20% w/v Polyethylene glycol 3,350	36.	7.3
37.	0.2 M Potassium sodium tartrate tetrahydrate	37.	20% w/v Polyethylene glycol 3,350	37.	7.4
38.	0.2 M Ammonium tartrate dibasic	38.	20% w/v Polyethylene glycol 3,350	38.	6.6
39.	0.2 M Sodium phosphate monobasic monohydrate	39.	20% w/v Polyethylene glycol 3,350	39.	4.7
40.	0.2 M Sodium phosphate dibasic dihydrate	40.	20% w/v Polyethylene glycol 3,350	40.	9.1
41.	0.2 M Potassium phosphate monobasic	41.	20% w/v Polyethylene glycol 3,350	41.	4.8
42.	0.2 M Potassium phosphate dibasic	42.	20% w/v Polyethylene glycol 3,350	42.	9.2
43.	0.2 M Ammonium phosphate monobasic	43.	20% w/v Polyethylene glycol 3,350	43.	4.6
44.	0.2 M Ammonium phosphate dibasic	44.	20% w/v Polyethylene glycol 3,350	44.	8.0
45.	0.2 M Lithium citrate tribasic tetrahydrate	45.	20% w/v Polyethylene glycol 3,350	45.	8.4
46.	0.2 M Sodium citrate tribasic dihydrate	46.	20% w/v Polyethylene glycol 3,350	46.	8.3
47.	0.2 M Potassium citrate tribasic monohydrate	47.	20% w/v Polyethylene glycol 3,350	47.	8.3
48.	0.2 M Ammonium citrate dibasic	48.	20% w/v Polyethylene glycol 3,350	48.	5.1



◊ Measured pH at 25 ° C

PEG/Ion Screen contains forty-eight unique reagents. To determine the formulation of each reagent, simply read across the page.

34 Journey
Aliso Viejo, CA 92656-3317 U.S.A.
Tel: (949) 425-1321 • Fax: (949) 425-1611
E-mail: tech@hamrsl.com
Website: www.hamptonresearch.com

HAMPTON RESEARCH
Solutions for Crystal Growth

Tube Number	Salt	Tube Number	Buffer 1
A1.	1.8 M Sodium Acetate pH 7.0	A1.	0.1 M Bis-Tris Propane pH 7.0
A2.	2.8 M Sodium Acetate pH 7.0	A2.	0.1 M Bis-Tris Propane pH 7.0
A3.	1.5 M Ammonium Chloride	A3.	0.1 M Sodium Acetate pH 4.6
A4.	1.5 M Ammonium Chloride	A4.	0.1 M Bis-Tris Propane pH 7.0
A5.	1.5 M Ammonium Chloride	A5.	0.1 M Tris pH 8.5
A6.	3.5 M Ammonium Chloride	A6.	0.1 M Sodium Acetate pH 4.6
A7.	3.5 M Ammonium Chloride	A7.	0.1 M Bis-Tris Propane pH 7.0
A8.	3.5 M Ammonium Chloride	A8.	0.1 M Tris pH 8.5
A9.	2.2 M Sodium Chloride	A9.	0.1 M Sodium Acetate pH 4.6
A10.	2.2 M Sodium Chloride	A10.	0.1 M Bis-Tris Propane pH 7.0
A11.	2.2 M Sodium Chloride	A11.	0.1 M Tris pH 8.5
A12.	3.2 M Sodium Chloride	A12.	0.1 M Sodium Acetate pH 4.6
B1.	3.2 M Sodium Chloride	B1.	0.1 M Bis-Tris Propane pH 7.0
B2.	3.2 M Sodium Chloride	B2.	0.1 M Tris pH 8.5
B3.	1.0 M di-Ammonium hydrogen Citrate	B3.	0.1 M Sodium Acetate pH 4.6
B4.	1.8 M di-Ammonium hydrogen Citrate	B4.	0.1 M Sodium Acetate pH 4.6
B5.	1.0 M tri-Ammonium Citrate pH 7.0	B5.	0.1 M Bis-Tris Propane pH 7.0
B6.	2.0 M tri-Ammonium Citrate pH 7.0	B6.	0.1 M Bis-Tris Propane pH 7.0
B7.	0.7 M tri-Sodium Citrate dihydrate	B7.	0.1 M Bis-Tris Propane pH 7.0
B8.	0.7 M tri-Sodium Citrate dihydrate	B8.	0.1 M Tris pH 8.5
B9.	1.2 M tri-Sodium Citrate dihydrate	B9.	0.1 M Bis-Tris Propane pH 7.0
B10.	1.2 M tri-Sodium Citrate dihydrate	B10.	0.1 M Tris pH 8.5
B11.	0.4 M Magnesium Formate	B11.	0.1 M Sodium Acetate pH 4.6
B12.	0.4 M Magnesium Formate	B12.	0.1 M Bis-Tris Propane pH 7.0
C1.	0.4 M Magnesium Formate	C1.	0.1 M Tris pH 8.5
C2.	0.7 M Magnesium Formate	C2.	0.1 M Bis-Tris Propane pH 7.0
C3.	2.0 M Sodium Formate	C3.	0.1 M Sodium Acetate pH 4.6
C4.	2.0 M Sodium Formate	C4.	0.1 M Bis-Tris Propane pH 7.0
C5.	2.0 M Sodium Formate	C5.	0.1 M Tris pH 8.5
C6.	3.5 M Sodium Formate	C6.	0.1 M Sodium Acetate pH 4.6
C7.	3.5 M Sodium Formate	C7.	0.1 M Bis-Tris Propane pH 7.0
C8.	3.5 M Sodium Formate	C8.	0.1 M Tris pH 8.5
C9.	1.2 M DL-Malic Acid pH 7.0	C9.	0.1 M Bis-Tris Propane pH 7.0
C10.	2.2 M DL-Malic Acid pH 7.0	C10.	0.1 M Bis-Tris Propane pH 7.0
C11.	1.4 M Sodium Malonate pH 7.0	C11.	0.1 M Bis-Tris Propane pH 7.0
C12.	2.4 M Sodium Malonate pH 7.0	C12.	0.1 M Bis-Tris Propane pH 7.0
D1.	2.5 M Ammonium Nitrate	D1.	0.1 M Sodium Acetate pH 4.6
D2.	2.5 M Ammonium Nitrate	D2.	0.1 M Bis-Tris Propane pH 7.0
D3.	2.5 M Ammonium Nitrate	D3.	0.1 M Tris pH 8.5
D4.	6.0 M Ammonium Nitrate	D4.	0.1 M Sodium Acetate pH 4.6
D5.	6.0 M Ammonium Nitrate	D5.	0.1 M Bis-Tris Propane pH 7.0
D6.	6.0 M Ammonium Nitrate	D6.	0.1 M Tris pH 8.5
D7.	1.5 M Sodium Nitrate	D7.	0.1 M Sodium Acetate pH 4.6
D8.	1.5 M Sodium Nitrate	D8.	0.1 M Bis-Tris Propane pH 7.0
D9.	1.5 M Sodium Nitrate	D9.	0.1 M Tris pH 8.5
D10.	4.0 M Sodium Nitrate	D10.	0.1 M Sodium Acetate pH 4.6
D11.	4.0 M Sodium Nitrate	D11.	0.1 M Bis-Tris Propane pH 7.0
D12.	4.0 M Sodium Nitrate	D12.	0.1 M Tris pH 8.5

† *Buffer pH is that of a 1.0 M stock prior to dilution with other reagent components. pH with HCl or NaOH.*

SaltRx HT contains ninety-six unique reagents.
To determine the formulation of each reagent, simply read across the page.

27632 El Lazo Road, Suite 100
Laguna Niguel, CA 92677-3913 U.S.A.
Tel: (949) 425-1321 • Fax: (949) 425-1611
e-mail: tech@hrmail.com
Internet: www.hamptonresearch.com

HAMPTON RESEARCH
Solutions for Crystal Growth

I Bioinformatics analysis of GH95 and GH97 proteins, gel filtration graphs for CAH06598 and CAH09443 and PCR amplification conditions used in the cloning of BF3763 and BF0855 genes.

ExPASy Home page Site Map Search ExPASy Contact us Proteomics tools S

Search Swiss-Prot/TrEMBL for Go Clear

ProtParam CAH06598

User-provided sequence:

```
        10         20         30         40         50         60
MKIKLLLLLC CGLWSSCNSY DYCPVTPSES DLVFTGLARS WDEAMPLGNA TVGALVWQRD
        70         80         90        100        110        120
STLRLSLDRT DLWDLRPVDS LSGDNFRFSW VKEHIRQKNY LPVQKKLDWP YDMNPAPSKI
       130        140        150        160        170        180
PGAAIEFPLE QIGTPTQVRL YLNNALCEAD WADGTQMQTF VHATEPIGWF VFRNLKTPIE
       190        200        210        220        230        240
PSIITPVYNK TKPDGSLDPV SGQDLHRLGY QQGKVVREGN QITYHQKGYG DFSYDVTVCW
       250        260        270        280        290        300
KQEGETLYGT WSVTSSLSGE QASEKAEAAL QRGLKHDYQA HLEYWDKYWA QSSITLPDSV
       310        320        330        340        350        360
LQKQYQNEMY KFGSTTREHS YPISLQAVWT ADNGKLPPWK GDYHHDLNTQ LSYWPAYTGN
       370        380        390        400        410        420
HLTEGMGYLN TLWNQRDAYK RYTRRYFGTE GMNIPGVCTL TGEPMGGWIQ YSMSQTVAAW
       430        440        450        460        470        480
LAQHFYLQWK YSADRTFLKE RAYPFIKDVA IYLEQISEVT PEGVRKLEFS SSPEIFDNSL
       490        500        510        520        530        540
QAWFSDMTNY DLAMMHFLFK ATSELAHELN LADEAGHWAS LEAQLPDYDI DEEGCLTFAK
       550        560        570        580        590        600
GYPYKESHRH FSHAMAIHPL GLIDWSDGEK SQHIIRATLK RLDKVGPDYW TGYSYSWLAN
       610        620        630        640        650        660
MKARAFDGEG AAQALKTFAE CFCLKNTFHA NGDQTQSGKS RFTYRPFTLE GNFAFAAGIQ
       670        680        690        700        710        720
EMLLQSHTGV IRIFPAIPKE WKDVSFENLR AMGAFLVSAR MEGGEINRVR IYSEKGGMLK
       730        740        750
MARPGTLKPN KNYTLSGTDI LNIDTQAGEW IELNP
```

References and documentation are available.

new Please note the modified algorithm for extinction coefficient.

Number of amino acids: 755

Molecular weight: 86102.1

Theoretical pI: 5.81

Amino acid composition:

Ala (A)	56	7.4%
Arg (R)	32	4.2%
Asn (N)	30	4.0%
Asp (D)	43	5.7%
Cys (C)	10	1.3%
Gln (Q)	37	4.9%
Glu (E)	45	6.0%
Gly (G)	55	7.3%
His (H)	21	2.8%
Ile (I)	32	4.2%
Leu (L)	72	9.5%
Lys (K)	41	5.4%
Met (M)	19	2.5%
Phe (F)	31	4.1%

Pro	(P)	37	4.9%
Ser	(S)	52	6.9%
Thr	(T)	48	6.4%
Trp	(W)	26	3.4%
Tyr	(Y)	39	5.2%
Val	(V)	29	3.8%
Pyl	(O)	0	0.0%
Sec	(U)	0	0.0%

(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 88
Total number of positively charged residues (Arg + Lys): 73

Atomic composition:

Carbon	C	3887
Hydrogen	H	5847
Nitrogen	N	1027
Oxygen	O	1138
Sulfur	S	29

Formula: $C_{3887}H_{5847}N_{1027}O_{1138}S_{29}$
Total number of atoms: 11928

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 201735
Abs 0.1% (=1 g/l) 2.343, assuming ALL Cys residues appear as half cystines

Ext. coefficient 201110
Abs 0.1% (=1 g/l) 2.336, assuming NO Cys residues appear as half cystines

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 34.84
This classifies the protein as stable.

Aliphatic index: 72.28

Grand average of hydropathicity (GRAVY): -0.472

CENTERFO RBIOLOGI CALSEQU ENCEANA LYSIS CBS

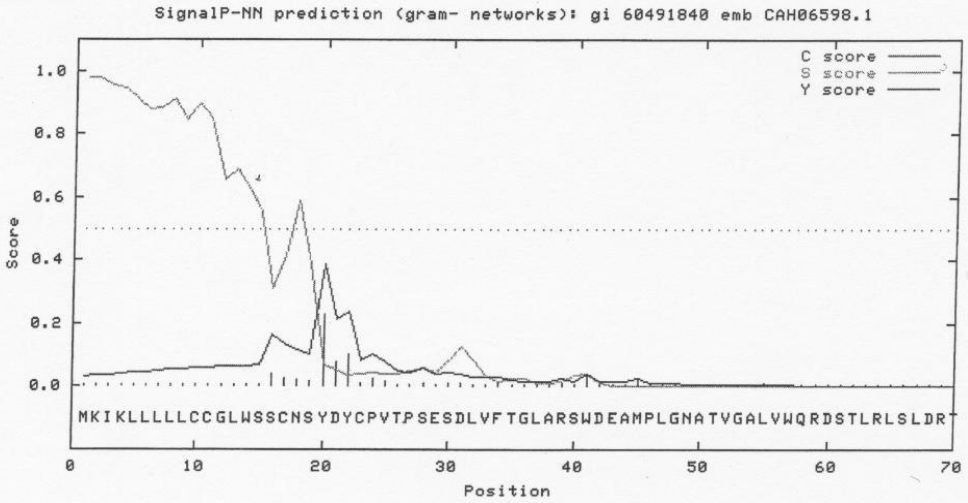
SignalP 3.0 Server - prediction results

Technical University of Denmark

Using neural networks (NN) and hidden Markov models (HMM) trained on Gram-negative bacteria

>gi_60491840_emb_CAH06598.1_ conserved hypothetical protein _Bacteroides fragilis NCTC 9343_

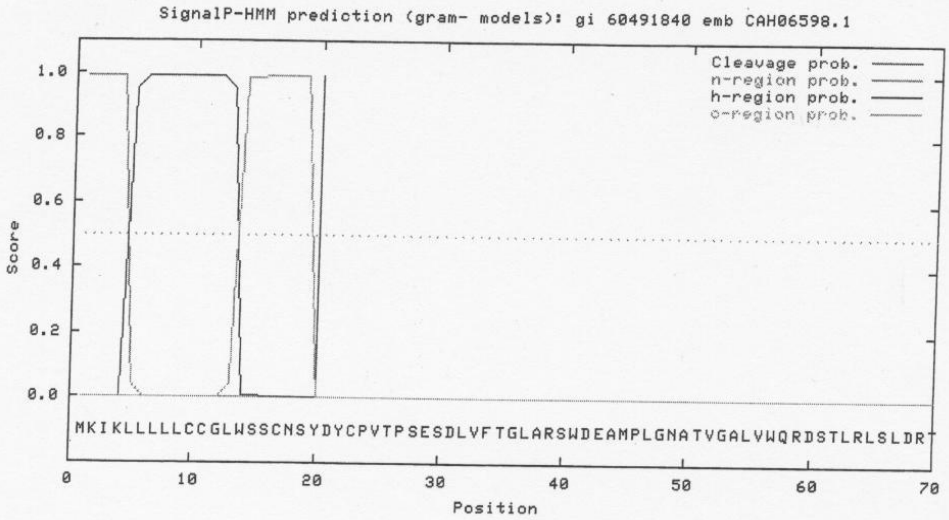
SignalP-NN result:



data

```
>gi_60491840_emb_CAH0  length = 70
# Measure  Position  Value  Cutoff  signal peptide?
  max. C    20       0.227  0.52    NO
  max. Y    20       0.386  0.33    YES
  max. S     2       0.982  0.92    YES
  mean S     1-19    0.751  0.49    YES
       D     1-19    0.568  0.44    YES
# Most likely cleavage site between pos. 19 and 20: CNS-YD
```

SignalP-HMM result:



data

```
>gi_60491840_emb_CAH06598.1_
Prediction: Signal peptide
Signal peptide probability: 0.991
Max cleavage site probability: 0.988 between pos. 19 and 20
```

gnuplot script

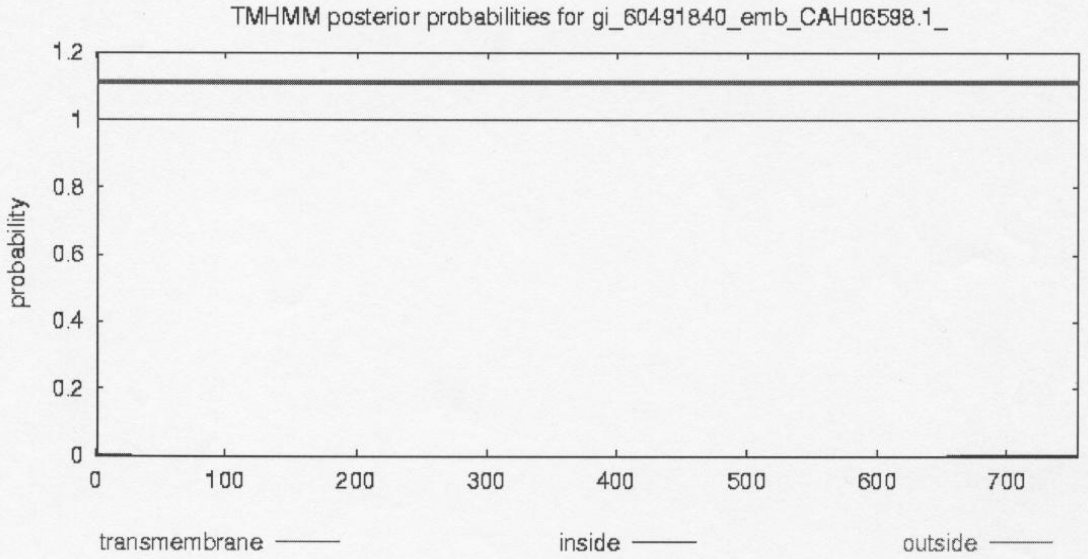
for making the plot(s)

Explain the output. Go **back**.

TMHMM result

HELP with output formats

```
# gi_60491840_emb_CAH06598.1_ Length: 755
# gi_60491840_emb_CAH06598.1_ Number of predicted TMHs:  0
# gi_60491840_emb_CAH06598.1_ Exp number of AAs in TMHs: 0.0179
# gi_60491840_emb_CAH06598.1_ Exp number, first 60 AAs:  0.01192
# gi_60491840_emb_CAH06598.1_ Total prob of N-in:        0.00063
gi_60491840_emb_CAH06598.1_     TMHMM2.0        outside      1   755
```



plot in postscript, script for making the plot in gnuplot, data for plot

---

ExPASy Home page Site Map Search ExPASy Contact us Proteomics tools Swiss-Prot

Search Swiss-Prot/TrEMBL for Go Clear

ProtParam CAH09443

User-provided sequence:

```
        10         20         30         40         50         60
MKRKMMSLLL ALAVISGSSV YAKVIDVMSP NGAIKVSVDI KDRIYYSVSY DNDQLLKDCY
        70         80         90        100        110        120
LNLQLQNETL GTNPHLRSTK RGTIDESVKR EIPFKNAIVR NHCNTLRMNF SGNYAVEFRV
       130        140        150        160        170        180
FDNGIAYRFV TDKKGDNIVM GEDFAINFPT NYKAHLSQPD GFKTSYECPY THVDTEKYAA
       190        200        210        220        230        240
TDRMSYLPVL IETDKAYKIL ISEADLSDYP CMFLKSTGKN GMQSIFPKAP LAFGEDGDRS
       250        260        270        280        290        300
LKITEEADYI AKTDGKRSFP WRMMVISKED KELIENEMVY NLSAPCVLED YSWIKPGQVS
       310        320        330        340        350        360
WEWWHDARLY GVDFRSGFNM DSYKYYIDFA SKFGIPYIIM DEGWAKNTRD PFTPNPTINL
       370        380        390        400        410        420
TELIKYGKDR NVKIVLWLPW LTVENHFDLF KTFADWGIAG VKIDFMDRSD QWMVNYYERV
       430        440        450        460        470        480
AKEAAKHKLF VDFHGAFKPA GLERKYPNVL SYEGVLGMEQ GGNCKPENSI YLPFMRNAVG
       490        500        510        520        530        540
PMDFTPGSMI SAQPEDNRST RANAMGSGTR AFQMALFIIF ESGLQMLADN PVYYYRELPC
       550        560        570        580        590        600
TEFITSVPVT WDETKVLYAK VGEAVVVAKR KGEQWFIGGI TGNQPQNIEI DLGFIPAGQS
       610        620        630        640
FTLTSFEDGI NADRQAMDYK KKESTVNNQT RMTLKMVRNG GWAGTIKMK
```

References and documentation are available.

new Please note the modified algorithm for extinction coefficient.

Number of amino acids: 649

Molecular weight: 74054.6

Theoretical pI: 6.33

Amino acid composition:

Ala (A)	42	6.5%
Arg (R)	29	4.5%
Asn (N)	37	5.7%
Asp (D)	44	6.8%
Cys (C)	7	1.1%
Gln (Q)	17	2.6%
Glu (E)	37	5.7%
Gly (G)	44	6.8%
His (H)	8	1.2%
Ile (I)	41	6.3%
Leu (L)	43	6.6%
Lys (K)	49	7.6%
Met (M)	27	4.2%
Phe (F)	34	5.2%
Pro (P)	30	4.6%
Ser (S)	39	6.0%
Thr (T)	35	5.4%
Trp (W)	13	2.0%

Tyr (Y)	33	5.1%
Val (V)	40	6.2%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 81
Total number of positively charged residues (Arg + Lys): 78

Atomic composition:

Carbon	C	3337
Hydrogen	H	5118
Nitrogen	N	868
Oxygen	O	973
Sulfur	S	34

Formula: $C_{3337}H_{5118}N_{868}O_{973}S_{34}$
Total number of atoms: 10330

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 121045
Abs 0.1% (=1 g/l) 1.635, assuming ALL Cys residues appear as half cystines

Ext. coefficient 120670
Abs 0.1% (=1 g/l) 1.629, assuming NO Cys residues appear as half cystines

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 32.76
This classifies the protein as stable.

Aliphatic index: 74.82

Grand average of hydropathicity (GRAVY): -0.370

CENTERFO RBIOLOGI CALSEQU ENCEANA LYSIS CBS

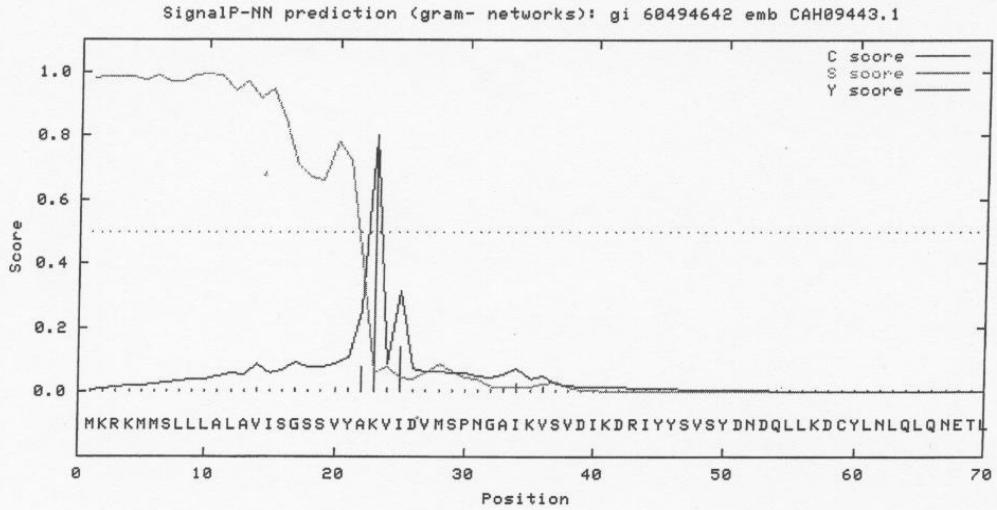
SignalP 3.0 Server - prediction results

Technical University of Denmark

Using neural networks (NN) and hidden Markov models (HMM) trained on Gram-negative bacteria

>gi_60494642_emb_CAH09443.1_ putative exported protein _Bacteroides fragilis NCTC 9343_

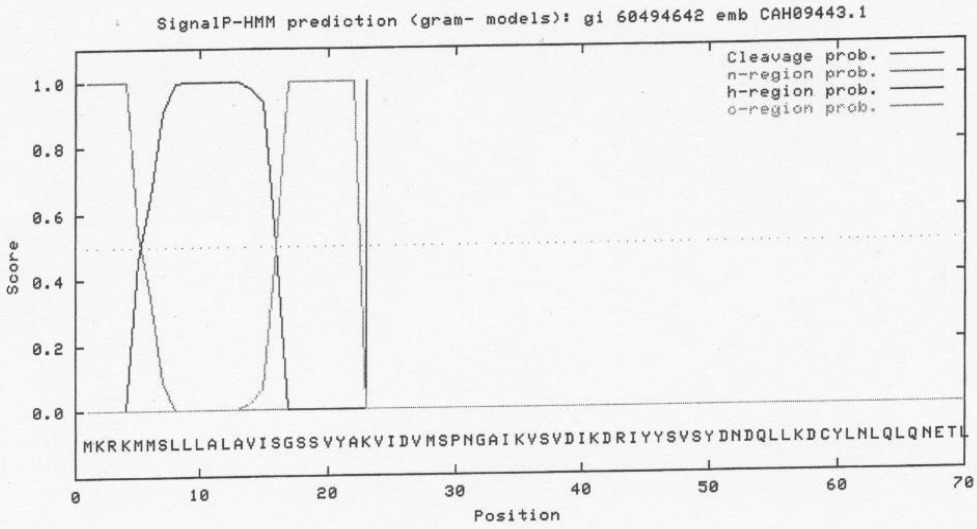
SignalP-NN result:



data

```
>gi_60494642_emb_CAH0  length = 70
# Measure  Position  Value  Cutoff  signal peptide?
  max. C    23       0.797   0.52   YES
  max. Y    23       0.800   0.33   YES
  max. S    10       0.993   0.92   YES
  mean S     1-22    0.882   0.49   YES
       D     1-22    0.841   0.44   YES
# Most likely cleavage site between pos. 22 and 23: VYA-KV
```

SignalP-HMM result:



data

>gi_60494642_emb_CAH09443.1_
Prediction: Signal peptide
Signal peptide probability: 1.000
Max cleavage site probability: 0.999 between pos. 22 and 23

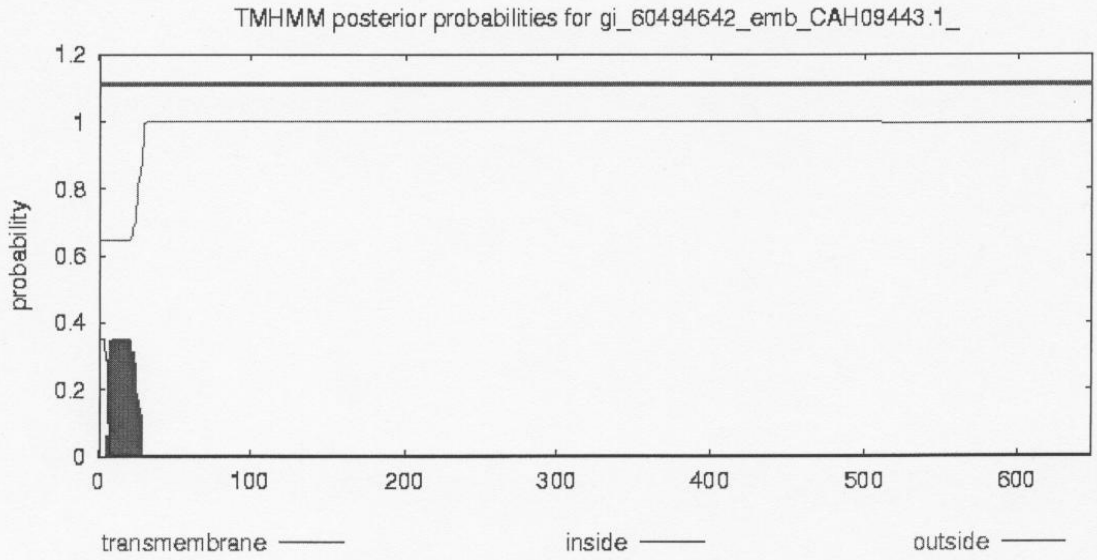
gnuplot script
for making the plot(s)

Explain the output. Go **back**.

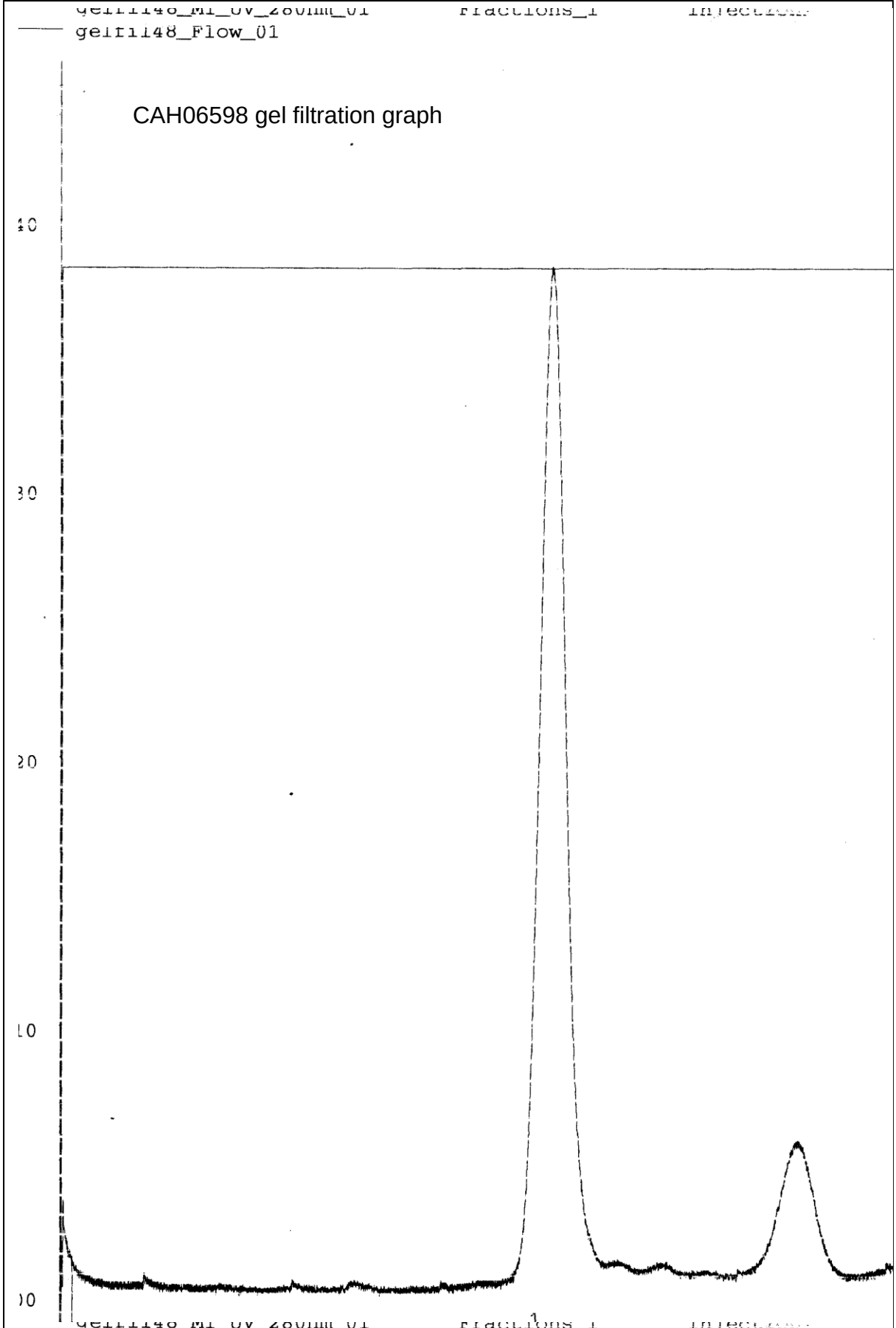
TMHMM result

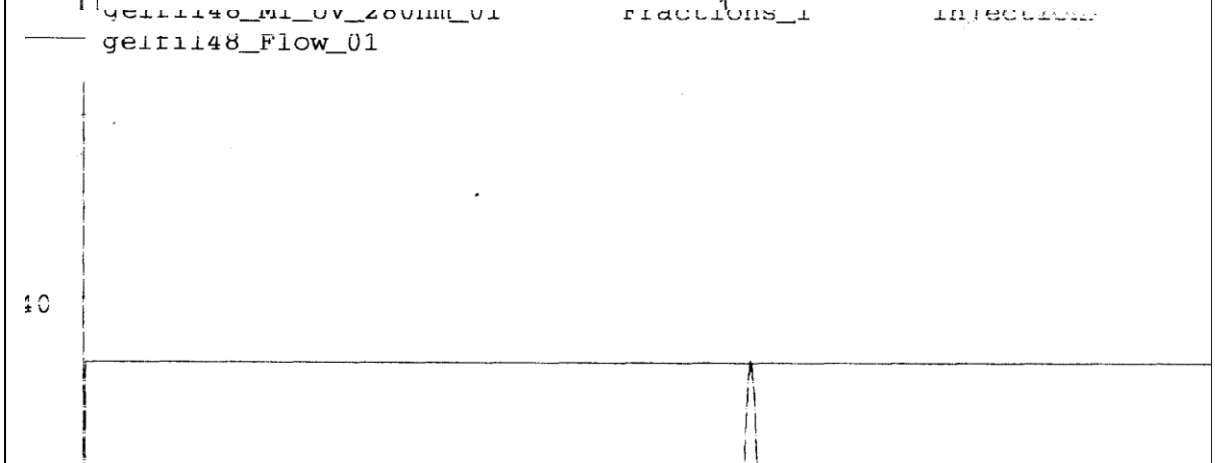
HELP with output formats

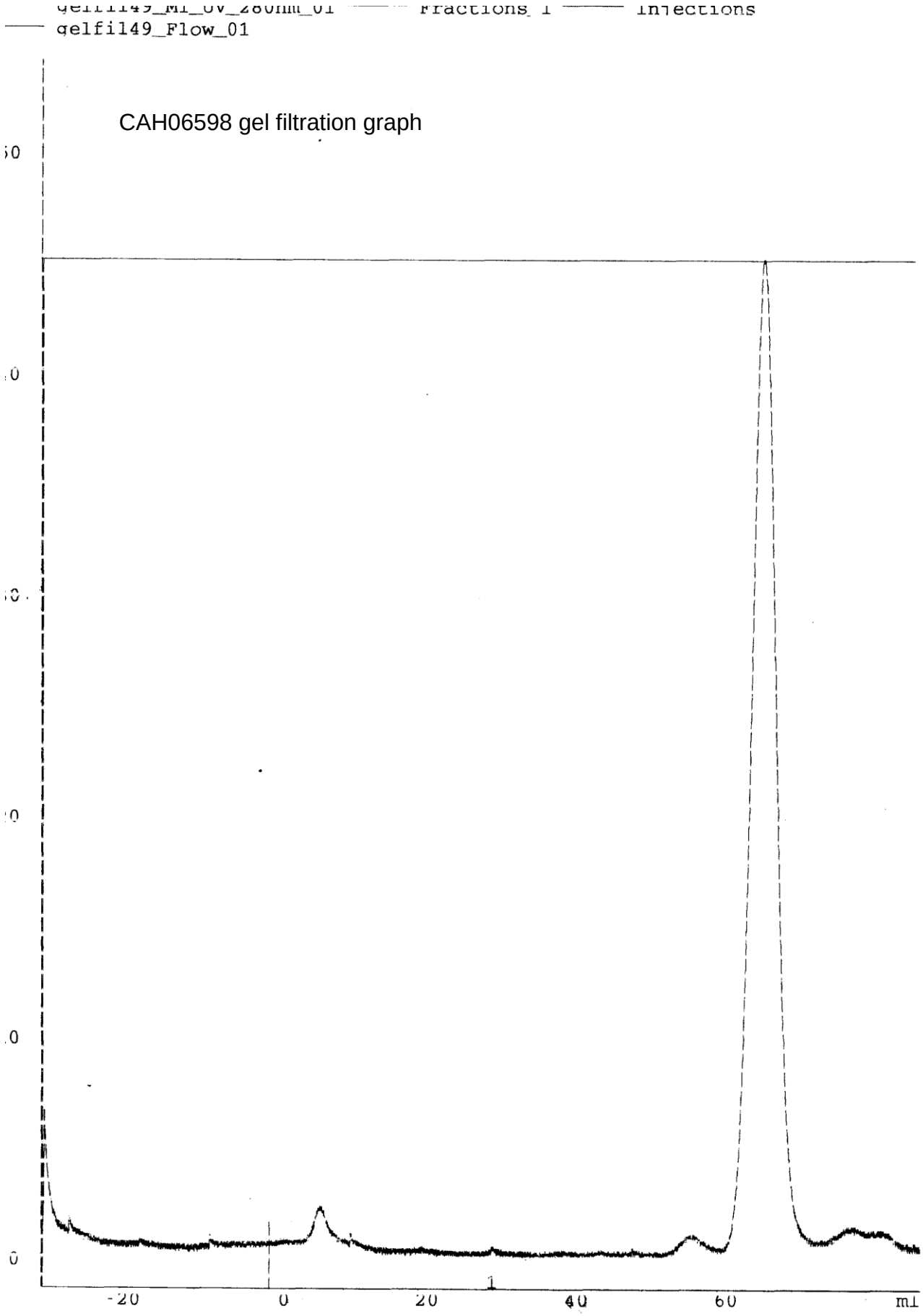
```
# gi_60494642_emb_CAH09443.1_ Length: 649
# gi_60494642_emb_CAH09443.1_ Number of predicted TMHs:  0
# gi_60494642_emb_CAH09443.1_ Exp number of AAs in TMHs: 7.31944000000000000000000000001
# gi_60494642_emb_CAH09443.1_ Exp number, first 60 AAs:  7.26249
# gi_60494642_emb_CAH09443.1_ Total prob of N-in:        0.35217
gi_60494642_emb_CAH09443.1_     TMHMM2.0        outside      1   649
```

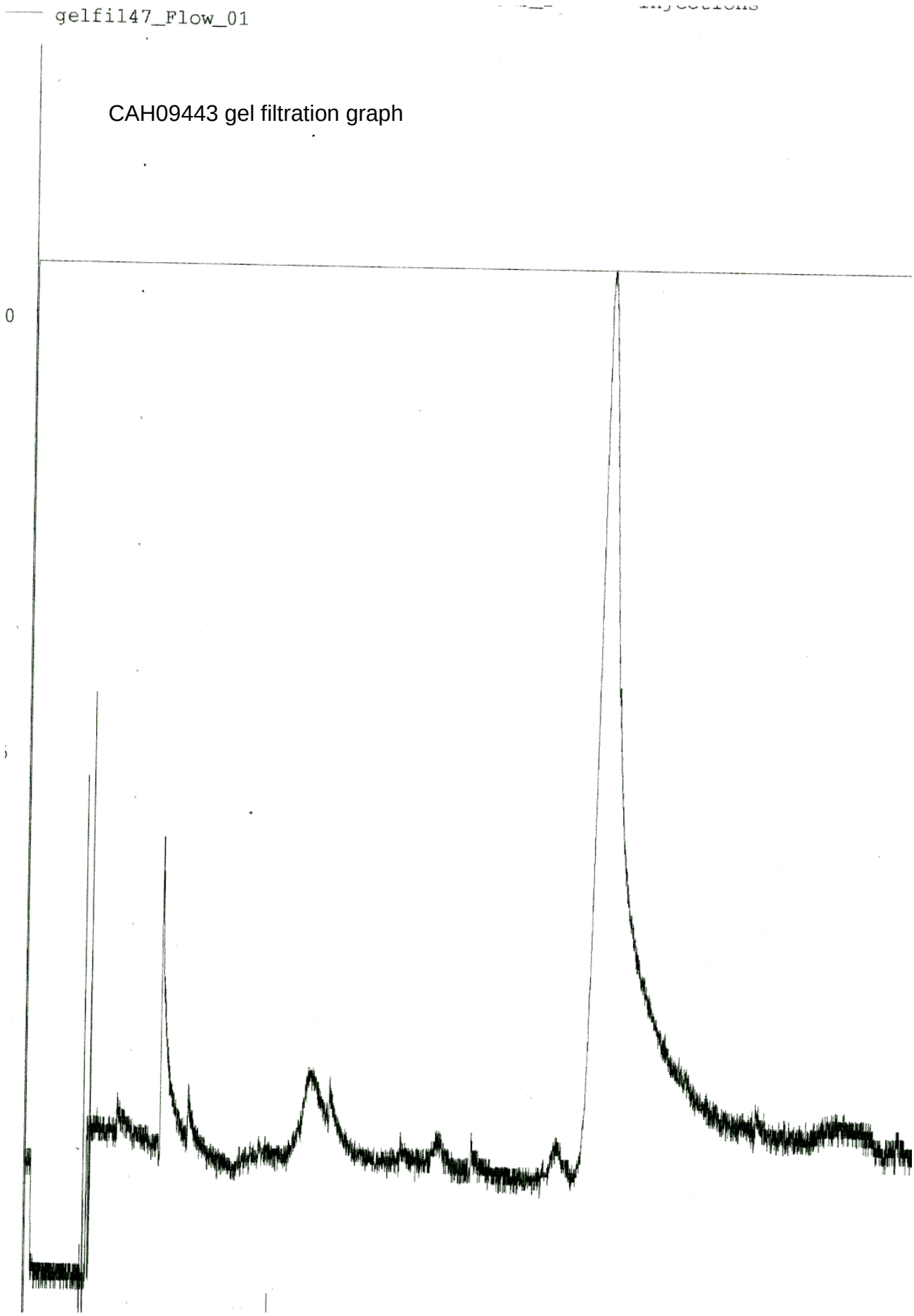


plot in postscript, script for making the plot in gnuplot, data for plot

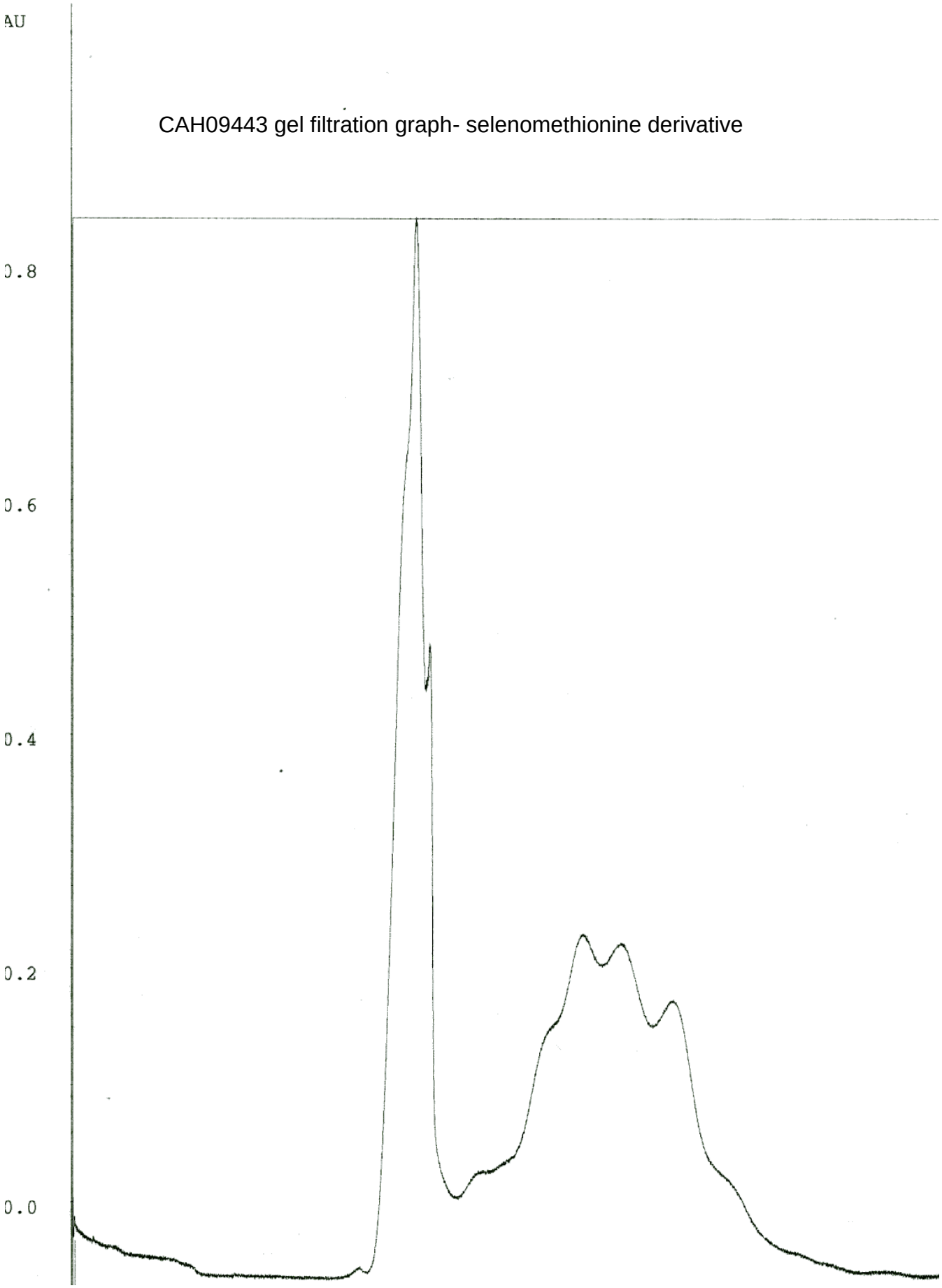

gelfi148_Flow_01
CAH06598 gel filtration graph
40
30
20
10
10


gelfi148_Flow_01
40


gelfil49_Flow_01
CAH06598 gel filtration graph
-20
0
20
40
60
ml


gelfil47_Flow_01
CAH09443 gel filtration graph
0

CAH09443 gel filtration graph- selenomethionine derivative



Step	Temperature	Time
Denature	94°C	2 min
Denature	94°C	15 s
Annealing	60.27°C	30 s
Extension	68°C	2 min 39 s
Repeated 4 more times		
Denature	94°C	15 s
Annealing	70.03°C	30 s
Extension	68°C	2 min 39 s
Repeated 24 more times		
Extension	68°C	10 min
Hold at 10°C		

PCR conditions for the amplification of the BF3763 gene from *B. fragilis*

Step	Temperature	Time
Denature	94°C	2 min
Denature	94°C	15 s
Annealing	60.27°C	30 s
Extension	68°C	2 min 39 s
Denature	94°C	15 s
Annealing	70.03°C	30 s
Extension	68°C	2 min 39 s
Extension	68°C	10 min
Hold at 10°C		

KOD polymerase based PCR amplification of BF3763 gene in *B. fragilis*

Step	Temperature	Time
Denature	94°C	2 min
Denature	94°C	15 s
	61 °C	30 s
Extension	72°C	1 min 53 s
	94°C	15 s
	80°C	30 s
Annealing	72°C	1 min 53 s
Denature	94°C	15 s
Annealing	73°C	30 s
Extension	72°C	1 min 53 s
Extension	72°C	10 min
Hold at 10°C		

PCR conditions for the amplification of the BF0855 gene from *B. fragilis*

Step	Temperature	Time
Denature	94°C	2 min
Denature	94°C	15 s
Annealing	61 °C	30 s
Extension	72°C	1 min 53 s
Denature	94°C	15 s
	80°C	30 s
Annealing	72°C	1 min 53 s
Denature	94°C	15 s
Annealing	73°C	30 s
Extension	72°C	1 min 53 s
Extension	72°C	10 min
Hold at 10°C		

Amplification of the BF3763 gene from transformed *E. coli* to screen for colonies containing successful inserts

Appendix J

Figure represents the standard curve for the Bradford assay